تليجرام : هنا سور الأزبكية
أكبر مكتبة رقمية

وزارة التعليم العالي والبحث
جامعة البصرة
كلية الإدارة والاقتصاد
قسم الإحصاء

تحليل الانحدار



$$F = k\frac{m_1 m_2}{r^2}$$

$$E(Y|X) = f(X, \beta)$$

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + u_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + u_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + u_n \end{aligned} \quad \ldots \quad (4-2)$$

تأليف

الأستاذ الدكتورة
زهرة حسن عباس التميمي

الأستاذ المساعد الدكتورة
فوزية غالب عمر السعدون

المدرس
ساهرة حسين زين الثعلبي

١٤٣٥ هـ ـ ٢٠١٤ م

وزارة التعليم العالي والبحث العلمي

جامعة البصرة

كلية الإدارة والاقتصاد

قسم الإحصاء

تحليل الانحدار



تأليف

الأستاذ الدكتورة
زهرة حسن عباس التميمي

الأستاذ المساعد الدكتورة
فوزية غالب عمر السعدون

الاستاذ المساعد
ساهرة حسين زين الثعلبي

١٤٣٥ هـ - ٢٠١٤ م

المقدمة:

تعد مادة تحليل الانحدار ، سواء أكانت كتاباً منهجياً أم بحوثاً تجريبية مهمة وضرورية لتقديم الصيغ العلمية والرياضية بصيغتها الجبرية، كأداة مناسبة من أدوات التحليل العلمي.

وقد أتيح لي تدريس مادة " تحليل الانحدار " لسنوات في جامعة البصرة على مستوى طلبة الدراسات العليا، فضلاً عن طلبة المراحل المنتهية في قسم الإحصاء والاقتصاد، فكان هذا الكتاب امتداداً لتلك السنوات الدراسية ومعززاً لها.

ولقد حاولت خلال الفصول المخصصة لهذه المادة، إعطاء فكرة شاملة ومبسطة قدر الإمكان عن مادة " الانحدار " التي تدرس في كبريات الجامعات والمعاهد في مختلف دول العالم. ويتميز هذا الكتاب باحتوائه على كم كبير من الأمثلة التي تم من خلالها تطوير المفاهيم العامة والخاصة التي تتضمنها أجزاؤه المختلفة، مع تطبيق البرنامج الإحصائي"" SPSS في بعض حلول تلك الأمثلة.

لقد شكلت فصول الكتاب طيفاً واسعاً إذ بالإمكان عدّه احد المراجع الجامعية الرئيسية في هذا الاختصاص إذ عرضت المسائل بشكل أكاديمي هادف. وعلى هذا الأساس تم تقسيم الكتاب إلى ثلاثة أجزاء تغطي مفردات مادة تحليل الانحدار المقرة للصفوف الثالثة إحصاء في كليات الإدارة والاقتصاد .

في الجزء الأول يتم دراسة نماذج الانحدار الخطي من خلال دراسة مشكلات الاستدلال الإحصائي كالتقدير واختبار الفرضيات فضلاً عن مشكلة التنبؤ بافتراض تحقق فروض التحليل.

أما الجزء الثاني فيتمثل بدراسة المشكلات الناجمة عن عدم تحقق فرضيات التحليل.

وفي الجزء الثالث تمت دراسة موضوعات أخرى في تحليل الانحدار ومنها طرائق اختيار أحسن المعادلات فضلاً عن المتغيرات الوهمية.

لقد حاولت قراءة مصادر عربية وأجنبية عديدة بعناية ودقة ثم كتابة الفصول دون الاقتباس مباشرة من أي من تلك المصادر(عدا بعض الأمثلة الافتراضية)، فقد حاولت توضيح التحليل بأمثلة قدر الإمكان.وقد ساعدني في ذلك الزميلتان الأستاذ المساعد د. فوزية غالب عمر والمدرس ساهرة حسين زين .

والآن وقد صار الكتاب بين أيدي القراء، نرجو من زملائنا وطلبتنا وغيرهم أن يقرؤوا الكتاب بعين ناقدة فاحصة، وليضعوا أصابعهم على مواطن الخلل والضعف، ويحيطوا كاتبه علماً بذلك، لعله يستطيع أن يسير به خطوة أخرى نحو الكمال، وليس هو ببالغه بغير ذلك.وليكن التواصل عبر البريد الالكتروني.

في الختام اتقدم بالشكر والتقدير لكل من قدم يد العون لانجاز هذا الكتاب، وهم كثر، لهم معزة خاصة في نفسي .

والله ولي التوفيق

الأستاذ الدكتورة زهرة حسن عباس التميمي

drzahra50@YAHOO.COM

مفـــردات الكتاب



أهم جروبات علي تليجرام
باحثون
هنا سور الازبكية
غواص في بحر الكتب
قناة مصر الثقافية و الفنية

الفصل الأول
طبيعة تحليل الانحدار

(١-١) الأساس التاريخي لمصطلح الانحدار Historical Origin of the Term (Regression)

تم تقديم الانحدار من قبل فرانسيس كالتون (Galton 1886) في مقالته[*] التي درس فيها استقرارية توزيع الأطوال في المجتمع وذلك باستخدام عينة لأكثر من ألف عائلة. وقد أكدت نتائجه على الرغم من وجود ميل لجميع الآباء طويلي القامة أن يحصلوا على أطفال طويلي القامة، وان الآباء قصار القامة لهم ميل للحصول على أطفال قصار القامة. فان متوسط طول الأطفال المولودين لآباء من طول معين يتحرك باتجاه (ينحدر) متوسط أطوال الأطفال في المجتمع ككل. وقد استخدم كالتون مصطلح الانحدار للإشارة إلى اتجاه الأطوال نحو المتوسط العام.

وقد تم تثبيت قانون كالتون من قبل كارل بيرسن (Karl Pearson 1903)[1] إذ جمع أكثر من ألف سجل بوصفها مجاميع لأطوال الآباء ، وتوصلت النتائج إلى أن متوسط الطول للأبناء عند تحديد مجاميع الآباء طوال القامة، يكون أقل من أطوال آبائهم . كما أن متوسط الطول للأبناء عند تحديد مجاميع الآباء قصار القامة يكون أطول من أطوال الآباء. وعليه تمت صياغة القانون على وفق الآتي: إنّ طول الأبناء ينحدر اعتماداً على متوسط طول الآباء.

وقد توالت وتعددت استخدامات هذا النوع من التحليل وشملت مختلف جوانب الحياة.

(١-٢) المفهوم الحديث للانحدار The modern interpretation of regression

يرتبط تحليل الانحدار بدراسة الاعتمادية لمتغير معين (يسمى المتغير المعتمد)، على متغير (أو متغيرات أخرى) تسمى (المتغيرات التوضيحية) بهدف الحصول على تقديرات ، والتنبؤ بمتوسط المجتمع للمتغير المعتمد بدلالة قيم معلومة (ثابتة) للمتغير(أو المتغيرات) التوضيحية بتكرار العينة.

وببساطة أن تحليل الانحدار هو وسيلة إحصائية تستخدم لتحليل البيانات التي تحتوي على متغيرين فأكثر عندما يكون الهدف هو اكتشاف طبيعة هذه العلاقة. ويُعد تحليل الانحدار من أكثر الطرائق الإحصائية استعمالاً في مختلف العلوم لأنه يصف العلاقة بين المتغيرات على هيأة معادلة.

---

* Frencis Galton , " Law of universal Regression ", 1886.

عالم الوراثة البريطاني (Sir Galton)

(1) K. Pearson and A. Lee, " On the laws of Inheritance ", Biometrica, vol.2, Nov.1903, pp(357-462) .

(١-٣) أمثلة عملية لتوضيح مفهوم الانحدار Examples

(١) قانون كالتون Galton الذي يهتم بدراسة الكيفية التي تتغير بها أطوال الأبناء في المتوسط مع توفر أطوال آبائهم. فللتنبؤ بمتوسط طول الأبناء مع معرفة طول الأب يستخدم طول الأب بوصفه المتغير التوضيحي (المتغير المستقل) في حين يكون طول الابن هو المتغير المعتمد.
وهناك أمثلة في كثير من النواحي الاقتصادية.

(٢) ومن الأمثلة الاقتصادية والإدارية:
أ. دراسة الإنفاق الاستهلاكي الشخصي واعتماده على الدخل المتاح الشخصي. فيكون الإنفاق الاستهلاكي الشخصي هو المتغير المعتمد وان الدخل المتاح الشخصي الحقيقي هو المتغير المستقل. وتفيد هذه الدراسة في تحديد الميل الحدي للاستهلاك ومتوسط التغيرات في الإنفاق الاستهلاكي على وفق التغيرات في الدخل الحقيقي.
ب. دراسة ناتج محصول معين واعتماده على درجة الحرارة، ومعدل الأمطار، وكمية ضوء الشمس والسماد، وهذه العلاقة تمكن من التنبؤ بمتوسط الناتج إذا توافرت معلومات عن المتغيرات التوضيحية :(درجة الحرارة، ومعدل سقوط الأمطار، وكمية ضوء الشمس والسماد ..)
ج. سعر سلعة معينة وتكاليف النقل لهذه السلعة، اذ يحدد سعر السلعة متغيراً معتمداً على تكاليف نقل هذه السلعة الذي يُعد متغيراً توضيحياً فيمكن تحديد التغيرات في سعر السلعة مع التغيرات لمتوسط تكاليف نقلها.
د. مستوى الأداء الوظيفي واعتماده على المؤهل الأكاديمي أو سنوات الخبرة لدى الموظف،
كما تستخدم في العلوم السلوكية.

(٣) العلاقة بين وزن الطفل وعمره، يمكن دراستها عن طريق تحليل الانحدار، اذ يتحدد وزن الطفل بعمره فيكون وزن الطفل هو المتغير المعتمد، وعمر الطفل يكون المتغير التوضيحي (المستقل).

(٤) دراسة أثر سرعة السيارة في عدد الحوادث.

(١-٤) طبيعة العلاقة بين المتغيرات Nature of the relation between variables

يمكن تحديد طبيعة العلاقة بين المتغيرات على وفق الصيغ التالية:

١- العلاقة المحددة Deterministic

الظواهر الطبيعية (أ) مثل قانون الجاذبية لنيوتن. فان كل جسيم في الأرض ينجذب إلى جسيمات أخرى على وفق قانون الجاذبية بقوة تتناسب طردياً مع حاصل ضرب كتلتيهما وعكسياً مع مربع المسافة بينهما:

$$F = k\frac{m_1 m_2}{r^2}$$

k : معامل التناسب.

m_i : كتلة الجسيم i ، i = 1,2

r : المسافة بين الجسيمين.

(ب) قانون اوم: التيار الكهربائي يتناسب مع مقدار الفولتية.

(ج) قانون بويل للغازات.

(د) قانون نيوتن للحركة.

٢- العلاقة شبه المحددة Semi deterministic

مثال: التكلفة الكلية تمثل التكلفة الثابتة يضاف إليها التكلفة المتغيرة، وان التكلفة المتغيرة تحدد بعدد وحدات الإنتاج لذلك فان التكلفة الكلية ليست صيغة محددة، فلابد من وجود جزء ثالث لمعالجة التوقف العرضي لإن الكميات المنتجة تتأثر بفترة العطل وبتكاليف تصليح العطلات العرضية فضلاً عن التغيرات في نوعية المواد الخام.

٣- العلاقة التجريبية Empirical relation

وتكون هذه العلاقة غير مسيطر عليها بقانون طبيعي او صيغة رياضية محددة.

مثال: محصول إنتاج البرتقال في تجربة زراعية معينة، فالعلاقة بين المحصول والسماد لا تتبع صيغة رياضية دقيقة لأن هناك عوامل عديدة غير السماد تؤثر في المحصول منها طبيعة التربة، ومستوى الري، والآفات الزراعية وغيرها.

ونماذج تحليل الانحدار ترتبط بالعلاقات الإحصائية وليست المحددة أو التامة. وفي العلاقات الإحصائية يتم التعامل مع المتغيرات العشوائية Random (stochastic) . وهذا بدوره يجعل القيم التنبؤية غير تامة بسبب وجود الخطأ (Error). بمعنى أن المتغير المعتمد يتضمن تغيرات عشوائية لا يمكن تفسيرها بالكامل برغم عدد المتغيرات التوضيحية المستخدمة.

(١-٥) الانحدار والسببية (Regression vs. Causation)

على الرغم من أن تحليل الانحدار يتعامل مع الاعتمادية بين أحد المتغيرات ومتغيرات أخرى تحدد مشكلة معينة، إلا أن ذلك لا يعني بالإطلاق استنتاج اتجاه السببية. ولقد أكد كاندال وستيوارت[1] أن العلاقات الإحصائية مهما كانت قوية لا يمكنها تحديد الارتباط ألسببي لان فكرة السببية تأتي من خارج الإحصاء وبشكل أكيد تعتمد على نظريات أخرى. فمثلاً في دراسة ناتج محصول معين، لا يوجد سبب إحصائي لافتراض أن معدل الأمطار لا يعتمد على ناتج المحصول. واختيارنا بجعل الناتج متغيراً معتمداً اعتماداً على افتراضات غير إحصائية. فالمنطق يقترح أن العلاقة لا يمكن أن تكون بالشكل العكسي، إذ لا يمكن السيطرة على معدل الأمطار من خلال التغيرات في الناتج، وعليه فان العلاقات الإحصائية في جميع أمثلة الانحدار لا يمكن أن تحدد السببية ، ولا بد من التركيز على الجوانب المسبقة والنظرية.

(١-٦) استخدامات تحليل الانحدار. Uses of regression analysis

يستخدم تحليل الانحدار لعدة أغراض أهمها:

١- وصف البيانات.

٢- تقدير المعلمات لإمكان الاستدلال على أهمية وقوة العلاقة بين المتغيرات.

٣- التنبؤ من خلال تقدير الاستجابة.

٤- السيطرة، إذ يمكن السيطرة على قيم المتغير المعتمد وذلك بتغير قيم المتغيرات التوضيحية.

(١-٧) أنواع الانحدار Types of regression

في فقرة سابقة تم توضيح مفهوم الانحدار بأنه وسيلة إحصائية لدراسة الاعتمادية بين متغير معين يسمى المتغير المعتمد وبين متغير آخر أو متغيرات أخرى تسمى المتغيرات التوضيحية، وعليه يمكن تقسيم الانحدار الى نوعين هما:

أ- الانحدار البسيط Simple regression

ب- الانحدار المتعدد General model or Multiple regression

فالانحدار البسيط يتضمن متغيراً معتمداً يرمز له (Y) ومتغيراً مستقلاً واحداً يرمز له (X) وتكون معادلة الانحدار:

$$Y = f(X, \beta)$$

حيث أن β معلمات غير معلومة وقد تكون (scalar) أو تكون بصيغة متجه (vector).

(1) M.G.Kendall & A. Stuart, "The Advanced theory of statistics ", charles Griffin publishers, New York, 1961, vol.2, ch.2, p.279.

كما يمكن ان نكتب كالآتي:

$$E(Y/X) = f(X, \beta)$$

اما في الانحدار المتعدد فيكون المتغير المعتمد (Y) موصوفاً بعدد من المتغيرات المستقلة (التوضيحية $X_k,......,X_2,X_1$. وتكون معادلة الانحدار :

$$Y = f(X_1, X_2, ..., X_k, \beta)$$

أو

$$Y = f(X_j, \beta_j) \qquad j = 1, ...k$$

ومن أجل إجراء الانحدار لابد من تحديد شكل الصيغة الدالية (f) التي يمكن استخدامها، ومن أبرز الصيغ الدالية هي الصيغة الخطية التي تتمثل بخط مستقيم، والصيغ غير الخطية وهي متعددة ومنها الصيغ متعددة الحدود (التربيعية والتكعيبية،....) والصيغ اللوغاريتمية والكسرية، واللوجستية، والمثلثية... الخ

وفي دراستنا هذه سيتم التركيز على الصيغ الخطية في المقام الأول مع التطرق إلى بعض الصيغ غير الخطية.

أسئلة الفصل الأول:

س١: عرف مفهوم تحليل الانحدار وبين أنواع الانحدار؟

س٢: ما أنواع العلاقات بين المتغيرات ؟ وما نوع العلاقة في تحليل الانحدار؟

س٣: اذكر أهم استخدامات تحليل الانحدار.

س٤: إعط أمثلة من الواقع العملي على تحليل الانحدار حسب أنواعها المختلفة ؟

س5: صحح الخطأ إن وجد في كلٍّ من العبارات التالية:

١- يقتصر تحليل الانحدار على وصف العلاقات شبه المحددة.

٢- إنّ المتغير المعتمد يتضمن تغيُّرات عشوائية والتي تم تفسيرها بالكامل بإدخال عددٍ من المتغيرات التوضيحية المستخدمة.

٣- إنّ إجراء الانحدار يتطلب صيغ دالية خطية.

٤- العلاقات التجريبية هي العلاقات التي تتم السيطرة عليها بقانون طبيعي.

٥- قانون كالتون هو علاقة انحدار لدراسة أثر أطوال الآباء في أطوال الأبناء باعتماد أطوال الآباء كمتغير معتمد.

٦- إنّ العلاقات الإحصائية تحدد الارتباط السببي.

الـفـصـل الـثـانـي
الانـحـدار الـخـطـي الـبـسـيـط Simple linear regression

(٢-١) مفهوم الانحدار الخطي البسيط .The concept of simple linear regression

نعني بالانحدار البسيط أن يكون المتغير المعتمد (Y) دالة بدلالة المعلمات β ومتغيراً توضيحياً واحداً هو المتغير المستقل (X) فقط.

وتكون الصيغة الدالية خطية بدلالة المعلمات وليس بالضرورة خطية بدلالة المتغير المستقل (X)، وبذلك تكون صيغة معادلة الانحدار الخطي البسيط على وفق الآتي:

$$\mathrm{E}(Y/X) = \beta_0 + \beta_1 X \qquad \text{(1-2)...}$$

β_1, β_0 معلمات ثابتة غير معلومة يمكن الحصول عليها بعد إجراء عملية الانحدار، وتسمى (β_0) المقطع الصادي أو الحد الثابت (Intercept)، وتمثل أيضا متوسط الاستجابة عندما X تساوي صفراً.

في حين (β_1) تمثل معلمة الميل (Slope coefficient).

وان $\mathrm{E}(Y/X)$ تشير إلى المتوسط الشرطي للمتغير Y .

والمعادلة (2-1) يطلق عليها " دالة الانحدار الخطي للمجتمع ".

وخدمة للهدف الرئيس لتحليل الانحدار وهو " تقدير معلمات العلاقة وتحقيق الاستدلال حول نتائج التقدير" يتم سحب عينة من المجتمع لكل من المتغيرين المعتمد (Y) والمستقل (X) ونفترض أن حجم العينة (n) وبذلك فان معادلة الانحدار هي:

$$\mathrm{E}(Y/X_i) = \beta_0 + \beta_1 X_i$$

وذلك يشير إلى قيم Y المشروطة للمشاهدة X_i ستتمحور حول المتوسط لقيم (Y_i) ولجميع العينات المسحوبة عند قيمة X_i. وعليه يمكن حساب الانحراف للقيم المنفردة (Y_i) حول قيمها المتوقعة وكالاتي:

$$\boldsymbol{u_i = Y_i - \mathrm{E}(Y/X_i)}$$

أو

$$Y_i = \mathrm{E}(Y/X_i) + u_i \qquad \ldots \qquad (2\text{-}2)$$

u_i : تمثل متغيراً عشوائياً غير مشاهد تأخذ قيماً موجبة او سالبة ويسمى أيضاً حد الخطأ العشوائي.(Stochastic error term) ويرمز له أيضاً (e_i) في بعض الكتب أو الدراسات.

والعلاقة (2-2) تتكون من جزأين.

الأول: $\mathrm{E}(Y/X_i)$ والذي يدل على متوسط (Y) الشرطي لجميع المشاهدات عند المستوى المحدد نفسه لـ (X). وتسمى هذه المركبة الجزء المحدد (deterministic) .

أما الثاني: u_i فهو الجزء العشوائي أو الاحتمالي وهي متغيرات تساعد على فهم التحليل من جانب، ومن جانب آخر تشكل الأساس لقياس دقة التقديرات وتعد بديلاً لجميع المتغيرات المحذوفة او المهملة التي قد تؤثر في سلوك المتغير المعتمد Y والتي لا يمكن تضمينها في معادلة الانحدار.

وبشكل عام فان معادلة الانحدار الخطي البسيط تكتب على وفق الآتي:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \ldots \qquad (3\text{-}2)$$

i = 1,2, . . . ,n

(٢-٢) تفسير معلمات النموذج The meaning of the parameters

(β_0) : متوسط الاستجابة عندما X = 0

(β_1) : تشير إلى مقدار التغير في Y عندما X_i تتغير بمقدار وحدة واحدة، وهي تمثل ميل الخط المستقيم وان إشارتها تدل على اتجاه العلاقة بين المتغيرين X , Y.

$$\beta_1 = \frac{dY}{dX}$$

واعتماداً على المعادلة (2-3) فان النموذج الخطي لأزواج المشاهدات يكتب كالآتي

$$\left.\begin{array}{l} Y_1 = \beta_0 + \beta_1 X_1 + u_1 \\ Y_2 = \beta_0 + \beta_1 X_2 + u_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_n + u_n \end{array}\right\} \qquad \ldots \qquad (4\text{-}2)$$

كما يمكن تمثيل المعادلات بالصيغة المصفوفية:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ . & . \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{pmatrix} \qquad \ldots \qquad (5-2)$$

أو:

$$Y = X\beta + u \qquad \ldots \qquad (6\text{-}2)$$

Y= متجه لمشاهدات المتغير المعتمد بترتيب .

X = مصفوفة بترتيب $(n \times 2)$ أعمدتها تمثل مشاهدات المتغير الوهمي الذي يعكس وجود المقطع الصادي وكذلك مشاهدات المتغير التوضيحي X .

u = متجه عمودي لمشاهدات المتغير العشوائي غير المشاهد.

(٢-٣) بناء نموذج انحدار خطي بسيط Construct simple linear model

لبناء نموذج انحدار خطي بسيط يتطلب إتباع الخطوات التالية:

١- تحديد المشكلة التي تتطلب دراستها، وباعتماد الأسس النظرية والمنطقية يحدد المتغير المعتمد (Y) وكذلك المتغير المستقل (X).

٢- جمع البيانات المطلوبة بحجم مناسب لحجم المجتمع الذي يمثل المشكلة.

٣- ترتيب البيانات على شكل أزواج مرتبة في جدول.

٤- التمثيل البياني للبيانات من خلال جعل المحور العمودي مخصصاً لقيم Y والمحور الأفقي يمثل قيم X للحصول على رسم الانتشار scatter diagram .

٥- من خلال رسم الانتشار تُحدد الصيغة الدالية المناسبة التي تكون خطية بدلالة المعلمات.

٦- استخدام الطرائق الإحصائية المناسبة لتقدير معلمات النموذج.

٧- الاستدلال حول صحة النتائج التقديرية.

٨- استخدام النتائج بعد التأكد من صحتها أو تعديلها قبل استخدامها لإغراض:
(اتخاذ القرارات او السيطرة او التنبؤ).

مثال (2-1) : نفترض أنّ الباحث يزمع دراسة إنتاج محصول البطاطا بدلالة المساحة المزروعة من ذلك المحصول.

١- المشكلة هي تحديد كمية الإنتاج من محصول البطاطا (Y) واعتماد المساحة المزروعة من المحصول متغيراً توضيحياً (X).

٢- جمع بيانات عن مزارع مختلفة لكل من المتغيرين Y و X بما يتلاءم مع حجم المجتمع المدروس، نفترض اختيار ١٠ مزارع عشوائياً.

٣- ترتيب البيانات في جدول (2-1)

جدول (2-1)

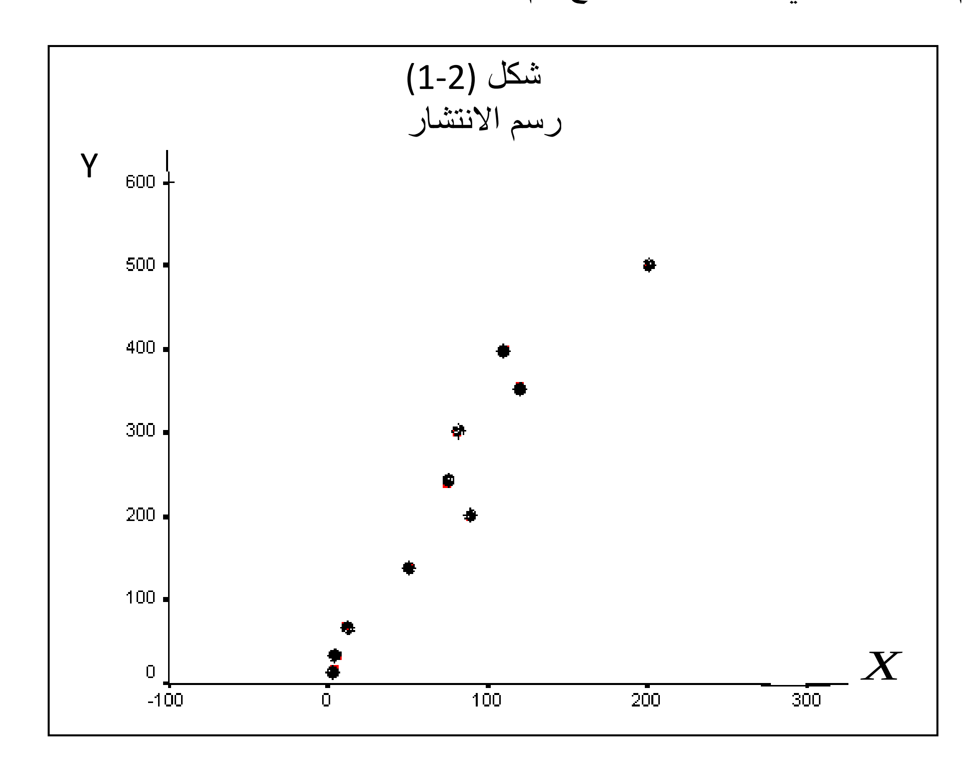
هكتار(X)	ألف كغم(Y)	المزرعة
٥٠	١٤٠	1
٢٠٠	٥٠٠	٢
١١٠	٤٠٠	٣
٨٠	٣٠٠	٤
١٢٠	٣٥٦	٥
٧٤.٥	٢٤٠.٥	٦
٨٨.٩	٢٠٠.٦	٧
٥.٧	٣٣.٥	٨
١١	٦٩.٨	٩
٣.٢	١٨.٧	١٠

743.32259.1 $\sum$

٤- التمثيل البياني للبيانات

نقسم المحور العمودي بشكل يتناسب مع قيم المتغير Y

نقسم المحور الأفقي بشكل يتناسب مع قيم المتغيرX



شكل (2-1)

رسم الانتشار

٥- يتضح من رسم الانتشار أن الصيغة الخطية ملائمة فيمكن استخدامها.

أي أن صيغة نموذج الانحدار المزمع استخدامه هو: $Y_i = \beta_0 + \beta_1 X_i + u_i$

i = 1,2, . . . ,10

٦- لتقدير المعلمات هناك طرائق إحصائية متعددة أهمها: المربعات الصغرى الاعتيادية وطريقة الإمكان الأعظم التي سيتم توضيحها في الفقرات التالية.

(4-2)تقدير دالة الانحدار الخطي البسيط. Estimating linear regression function

كما نوهنا سابقاً أن هدف تحليل الانحدار التركيز على تقدير دالة الانحدار للمجتمع بالاعتماد على عينة مختارة تمثل ذلك المجتمع، ويزخر التراث الإحصائي بطرائق متعددة لهذا الغرض غير أن أكثر هذه الطرائق استخداماً هي طريقة المربعات الصغرى الاعتيادية (OLS)، وسيتم في هذه الفقرة مناقشة هذه الطريقة.

تقدير معلمات الانحدار الخطي البسيط بطريقة المربعات الصغرى الاعتيادية.

(1-4-2)طريقة المربعات الصغرى الاعتيادية Ordinary least square(OLS) Method

يرمز لها اختصاراً (OLS) وهي أكثر الطرائق استخداماً في تقدير المعلمات. وتستند هذه الطريقة الى مبدأ ((تصغير مجموع مربعات الأخطاء)) فهي تسعى لإيجاد المعلمات التي تجعل مجموع مربعات الخطأ أقل ما يمكن.

بعبارة أخرى أن المعيار الذي تستند اليه هذه الطريقة هو تأكيد أن الخط الذي يتم تقديره للبيانات يتمثل بوجوب أن تكون مجموع مربعات المسافة العمودية لأي نقطة من البيانات عن ذلك الخط أصغر ما يمكن.

ويتم تربيع المسافة لمنع حذف المسافة الكبيرة الموجبة مع المسافات الكبيرة السالبة.

وهذا المعيار فعال جدا، وبذلك يكون الخط المستقيم المقدر هو ذلك الخط الذي يمر عبر متوسط البيانات (أي يتوسط البيانات للعينة). وان المقطع والميل الصادي المقدر لذلك المستقيم هما $\hat{\beta}_0, \hat{\beta}_1$.

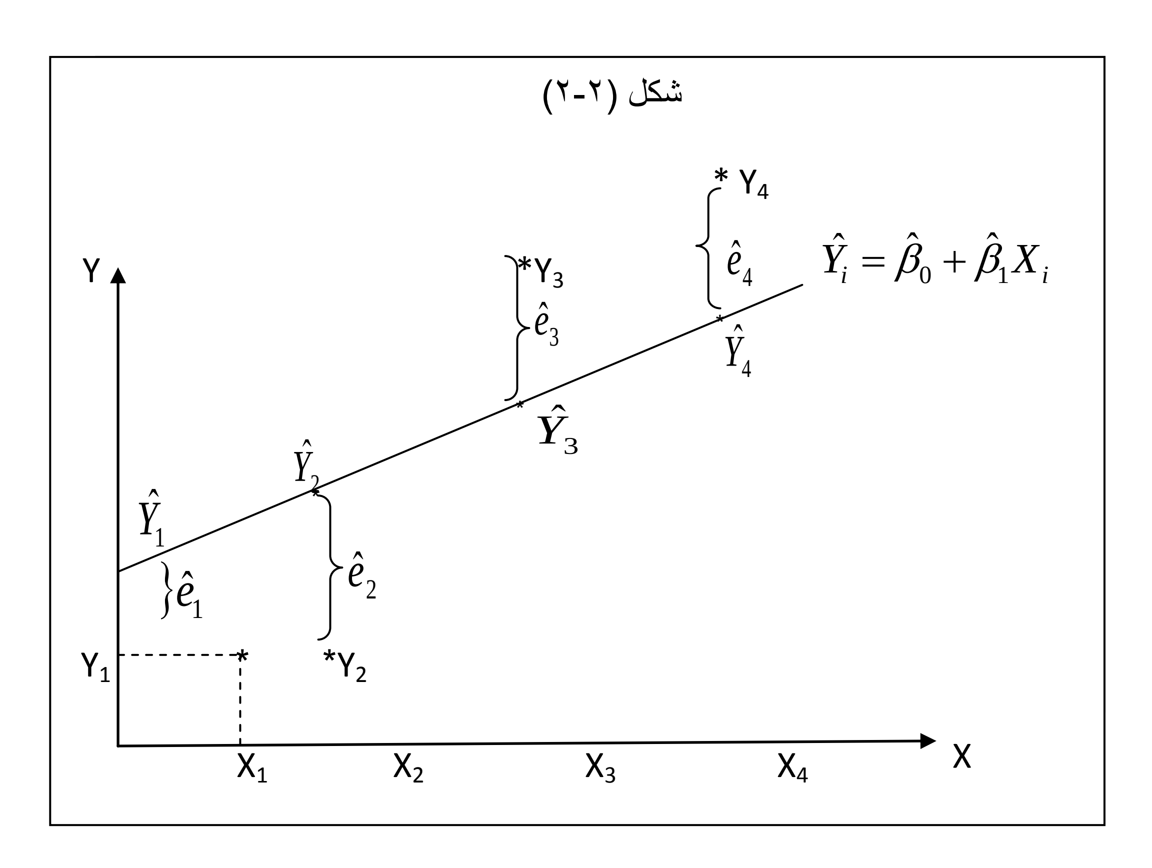
ويكون الخط المستقيم المقدر: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

وان المسافات العمودية من أي نقطة من نقاط العينة إلى المستقيم المقدر هي بواقي المربعات الصغرى ويرمز لها: $\hat{u}_i = Y_i - \hat{Y}_i$

أو:

$$e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

والبواقي موضحة في الشكل (٢-٢)



ويرمز لمجموع مربعات البواقي $S = \sum_{i=1}^{n} e_i^2$ S

$$S(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

والتي تمثل دالة الهدف التي تسعى الطريقة إلى تصغيرها. وهي دالة بدلالة المعلمات $\hat{\beta}_1, \hat{\beta}_0$.

وعليه فان التفاضل الجزئي لهذه الدالة بدلالة $\hat{\beta}_1, \hat{\beta}_0$ على التوالي يساوي صفراً يمثل الشرط الضروري[*](necessary condition).

$$\frac{\partial S}{\partial \hat{\beta}} = 0$$

$$= -2\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) \qquad ... \qquad (6-2)$$

(*) أما الشرط الكافي (Sufficient condition) وهو شرط التحقيق من صحة القيم الحرجة في تحقيق تصغير الدالة فيكون باستخدام المشتقات الجزئية الثانية التي يمكن وضعها في مصفوفة تسمى مصفوفة هيسين Hessian $H = (\frac{\partial^2 S}{\partial \beta_i^2})$

$$\frac{\partial S}{\partial \hat{\beta}_1} = 0$$

$$= -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) X_i \qquad \ldots \qquad (7-2)$$

وبإعادة ترتيب المعادلتين كالآتي:

$$n\,\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_i = \Sigma Y_i \qquad \ldots \qquad (8-2)$$

$$\hat{\beta}_0 \Sigma X_i + \hat{\beta}_1 \Sigma X_i^2 = \Sigma X_i Y_i \qquad \ldots \qquad (9-2)$$

وتسمى المعادلتان(2-8) و (2-9) بالمعادلات الطبيعية (Normal Equations) وبحل المعادلتين

نحصل على قيم $\hat{\beta}_1, \hat{\beta}_0$ المطلوبة في معادلة الانحدار

ويمكن إيجاد الحل أما باستخدام طريقة الحذف والتعويض. نضرب المعادلة (2-8) بـ (ΣX_i)

والمعادلة (2-9) بـ (n) ثم نطرح المعادلة (2-9) من (2-8) ثم ترتيب الحل للحصول على:

$$\hat{\beta}_1 = \frac{n\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \qquad \ldots \qquad (10-2)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \qquad \ldots \qquad (11-2)$$

حيث أن:

$$\bar{X} = \frac{\Sigma X_i}{n} \quad , \quad \bar{Y} = \frac{\Sigma Y_i}{n}$$

والمعادلات (2-10) و (2-11) تمثل قانون إيجاد المعلمات $\hat{\beta}_1, \hat{\beta}_0$ على وفق طريقة المربعات الصغرى بشكل عام ولأي عينة مختارة.

كما يمكن الحصول على الحل بالصيغة المصفوفية، اذ تتم كتابة المعادلات الطبيعية بالصيغة المصفوفية:

$$\begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix} \qquad \ldots \qquad (12-2)$$

أو :

$$X'X\hat{\beta} = X'Y \qquad \ldots \qquad (12-2)'$$

حيث أنّ:

$$Y_{(n\times1)}=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix}\quad,\quad X_{(n\times2)}=\begin{bmatrix}1&X_1\\1&X_2\\\vdots&\vdots\\1&X_n\end{bmatrix}\quad,\quad \hat{\beta}_{(2\times1)}=\begin{pmatrix}\hat{\beta}_0\\\hat{\beta}_1\end{pmatrix}$$

لذا فأنّ:

$$X'X=\begin{pmatrix}n&\Sigma X_i\\\Sigma X_i&\Sigma X_i^2\end{pmatrix}\quad,\quad X'Y=\begin{bmatrix}\Sigma Y_i\\\Sigma X_iY_i\end{bmatrix}$$

وبحل المعادلات الطبيعية بالصيغة المصفوفية. نضرب طرفي المعادلة '(12-2) بمعكوس $(X'X)$:

$$(X'X)^{-1}=\begin{bmatrix}\dfrac{\Sigma X_i^2}{nS_{XX}}&\dfrac{-\Sigma X_i}{nS_{XX}}\\\dfrac{-\Sigma X_i}{nS_{XX}}&\dfrac{n}{nS_{XX}}\end{bmatrix}$$

فنحصل على:

$$\hat{\beta}=(X'X)^{-1}X'Y$$

وبعد التبسيط:

$$\begin{pmatrix}\hat{\beta}_0\\\hat{\beta}_1\end{pmatrix}=\begin{bmatrix}\bar{Y}-\hat{\beta}_1\bar{X}\\\dfrac{S_{xy}}{S_{xx}}\end{bmatrix}$$

حيث أن:

$$S_{xx}=\Sigma X_i^2-\frac{(\Sigma X_i)^2}{n}$$

تمثل مجموع مربعات قيم X عن انحرافاتها[*].

---

* هناك صيغ متعددة لكتابة كلٍّ من S_{xx} و S_{xy}

$$S_{xx}=\Sigma(X_i-\bar{X})^2=\Sigma X_i^2-\frac{(\Sigma X_i)^2}{n}=\Sigma X_i^2-n\bar{X}^2$$

$$S_{xy}=\Sigma(X_i-\bar{X})(Y_i-\bar{Y})=\Sigma Y_i(X_i-\bar{X})=\Sigma X_i(Y_i-\bar{Y})=\Sigma X_iY_i-n\bar{X}\bar{Y}$$

و

$$S_{xy} = \Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}$$

تمثل مجموع حاصل ضرب انحرافات كل من X ، Y .

أي أنَّ

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{n\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2}$$

وهي المعادلات نفسها التي تم الحصول عليها في المعادلتين (2-10) و (2-11).

ويمكن الحصول على النتيجة نفسها باستخدام الحل بطريقة المحددات أو ما يسمى طريقة كريمر.

جدول (2-2)

جدول حسابات المجاميع

المشاهدات	Y	X	X^2	XY
1	١٤٠	٥٠	2500	7000
٢	٥٠٠	٢٠٠	40000	100000
٣	٤٠٠	١١٠	12100	44000
٤	٣٠٠	٨٠	6400	24000
٥	٣٥٦	١٢٠	14400	42720
٦	٢٤٠.٥	٧٤.٥	5550.25	17917.25
٧	٢٠٠.٦	٨٨.٩	7903.21	17833.34
٨	٣٣.٥	٥.٧	32.49	190.95
٩	٦٩.٨	١١	121	767.8
١٠	١٨.٧	٣.٢	10.24	59.84
Σ	2259.1	743.3	89017.19	254489.18

وبالرجوع إلى المثال باستخدام المربعات الصغرى الاعتيادية فان جدول حسابات المجاميع جدول (2-2) يساعد في تشكيل المعادلات الطبيعية بصيغة المصفوفات:

$$\begin{pmatrix} 10 & 743.3 \\ 743.3 & 89017.19 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 2259.1 \\ 254489.18 \end{pmatrix}$$

وان معادلة الانحدار المقدرة هي:

$$\hat{Y}_i = 35.35 + 2.564 X_i$$

تفسير المعْلمات:

زيادة المساحة المزروعة بمقدار هكتار واحد سوف يساعد على زيادة الإنتاج بمقدار (2.564) الف كغم. وان العلاقة بين المساحة المزروعة وبين الإنتاج علاقة طردية.

(2-4-2)فروض التحليل Assumption underlying the analysis

أن الهدف من تحليل الانحدار هو تقدير معلمات النموذج ثم الاستدلال حول قيم المعلمات الحقيقية. إذ يمتد عمل تحليل الانحدار لمعرفة كم تكون قيم المعلمات المقدرة $\hat{\beta}_1, \hat{\beta}_0$ قريبة من اقيامها في المجتمع أو يسعى التحليل لتحديد مدى قرب قيم Y المقدرة $(\hat{Y})$ لمتوسط الاستجابة الحقيقي $E(Y/X_i)$. وعليه فان تحليل الانحدار لا يكتفي بتحديد الصيغة للانحدار كما في العلاقة (٢-٢) ولكن لابد من عمل فرضيات محددة حول الطبيعة التي تتولد فيها قيم Y_i. اذ يتضح من العلاقة (٢-٢) بان Y_i تعتمد على قيم (X_i) وكذلك (u_i) ، ولذا لابد من التأكيد حول كيفية توليد قيم X_i و u_i ، ومن هنا فان الفرضيات الخاصة بمشاهدات المتغير X وكذلك قيم الأخطاء تكون غاية في الأهمية لإعطاء التفسير المناسب لمعلمات الانحدار. وسيتم في هذه الفقرة التركيز على الفرضيات الأساسية الخاصة بنموذج الانحدار التقليدي (CLRM) (classical linear regression model)

الفرضية ١:

نموذج الانحدار يكون خطياً بدلالة المعلمات وكما في العلاقة (3-2):

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \ldots \qquad (3\text{-}2)$$

ولابد من التأكيد ان العلاقة (3-2) ربما تكون غير خطية بدلالة المتغير المستقل X أو المتغير المعتمد Y ، فالعلاقات:

$$(1)\; Y_i = \beta_0 + \beta_1 X_i^2 + u_i$$

و

$$(2)\; Y_i = \beta_0 + \beta_1 \sqrt{X_i} + u_i$$

$$(3)\ \frac{1}{Y_i} = \beta_0 + \beta_1 X_i + u_i$$

$$(4)\ \log Y_i = \beta_0 + \beta_1 \log X_i + u_i$$

جميعها خطية بدلالة المعلمات، على الرغم من كونها غير خطية بدلالة المتغيرات Y أو X او كلاهما.

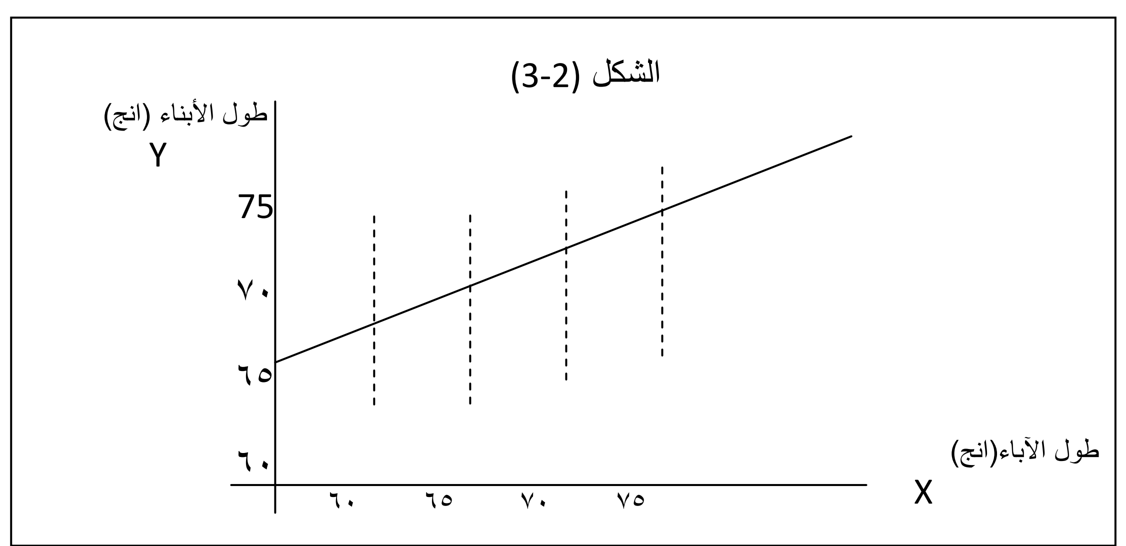
غير ان العلاقة: $Y = \beta_0 + \sqrt{\beta}_1 X_i + u_i$

هي علاقة غير خطية بدلالة المعلمة β_1 وعليه فلا تحقق الفرض ١.

الفرض ٢ :

قيم المتغير X ثابتة في العينات المتكررة وبعبارة أخرى ان المتغير X يفترض ان يكون غير عشوائي.

ففي مثال قانون كالتون للتنبؤ عن متوسط الطول للأبناء مع معرفة طول الآباء والذي يتم توضيحه في رسم الانتشار الشكل(2-3)



عند طول معين للآباء (X_i) مثلاً (70 انج) يتم سحب عائلة عشوائياً ويتم النظر بطول الأبناء Y مثلاً (٦٥ انج)، وتسحب عائلة أخرى عشوائياً وينظر بطول الأبناء Y مثلاً (٧٢ انج). ففي أي من السحبات (عينات متكررة) فان قيم X تبقى ثابتة عند (٧٠ انج). ويمكن اعادة هذه العملية لجميع قيم X المحددة للمجتمع الذي تم دراسته.

وهذا يعني تحليل الانحدار هو تحليل انحدار شرطي لقيم محددة من المتغير (X_i) .

الفرض ٣ :

المتغير العشوائي u له متوسط صفري عند قيم محدودة لـ X .

$$E(u_i / X_i) = 0$$

الشكل (2-3) يوضح، عند قيم معلومة من X_i فان u_i بعضها فوق الخط المقدر وبعضها الآخر تحت الخط المقدر وبذلك في المتوسط تكون انحرافات u_i عن الخط المقدر عند قيمة معينة من X يجب ان تساوي صفراً.

ولابد من التأكيد بان تحقق هذه الفرضية يعني ان العلاقة (٢-٢) متحققة.

أي :

$$\mathrm{E}(Y_i / X_i) = \beta_0 + \beta_1 X_i$$

الفرض ٤ :

تجانس تباين المتغير العشوائي (Homoscedasticity) u_i عند قيم محددة لـ X ، تباين المتغير العشوائي u_i يكون متساوياً لجميع المشاهدات.

أي:

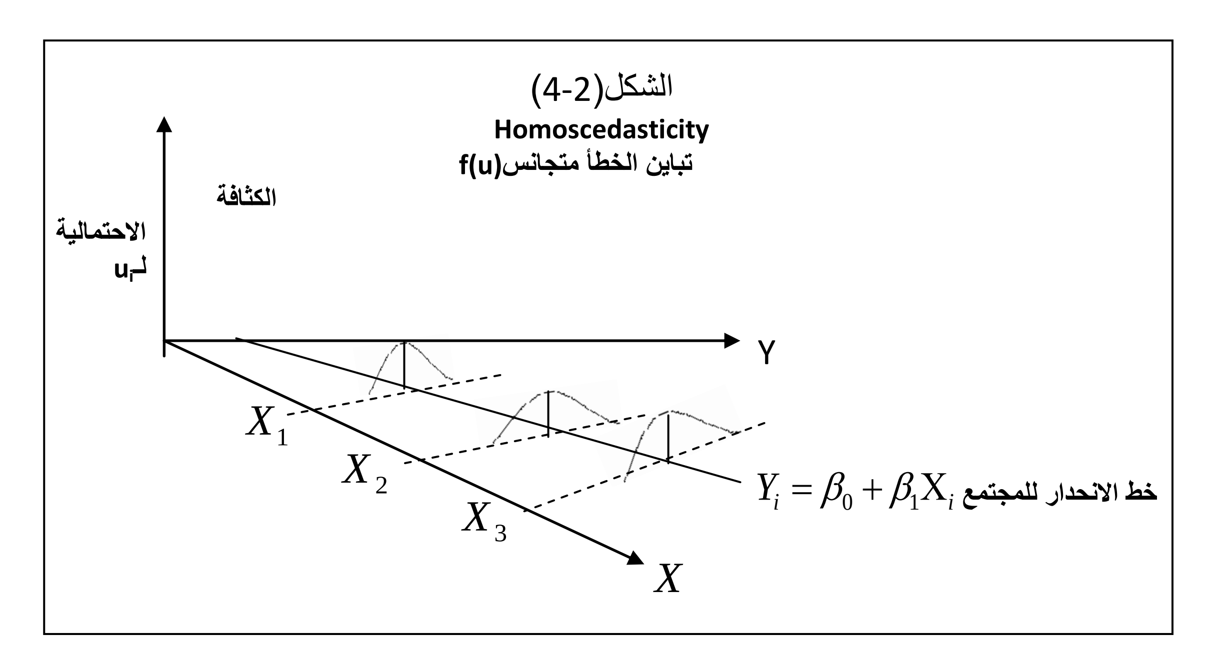
$$\begin{aligned} \mathrm{var}(u_i / X_i) &= \mathrm{E}(u_i - \mathrm{E}(u_i) / X_i)^2 \\ &= \mathrm{E}(u_i^2 / X_i) \\ &= \sigma^2 \qquad \forall \quad i = 1,2,\cdots,n \end{aligned}$$

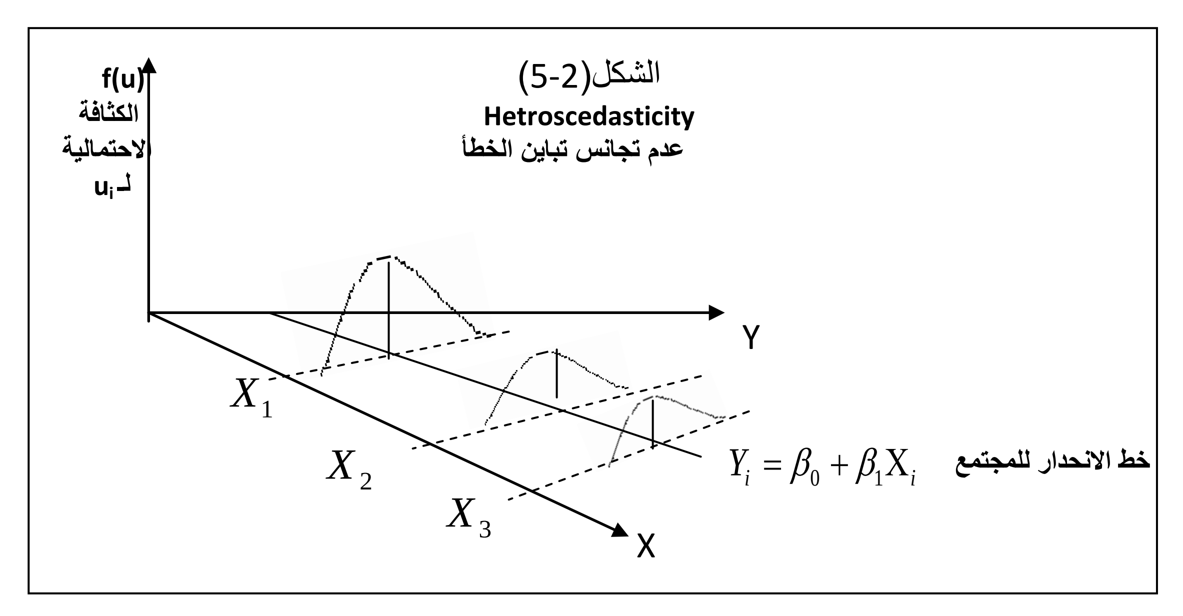
وعلى النقيض من هذه الفرضية فان التباين الشرطي لـ Y يتغير مع X ، وتسمى هذه الحالة عدم التجانس Hetroscedasticity .

وبالرموز:

$$\mathrm{var}(u_i / X_i) = \sigma_i^2$$

ويمكن توضيح الحالتين بالشكلين (2-4) و (2-5) .





الشكل (5-2) يوضح: $\text{var}(u/X_1) \langle \text{var}(u/X_2) \langle \text{var}(u/X_3)\cdots$

<u>الفرض (٥) :</u>

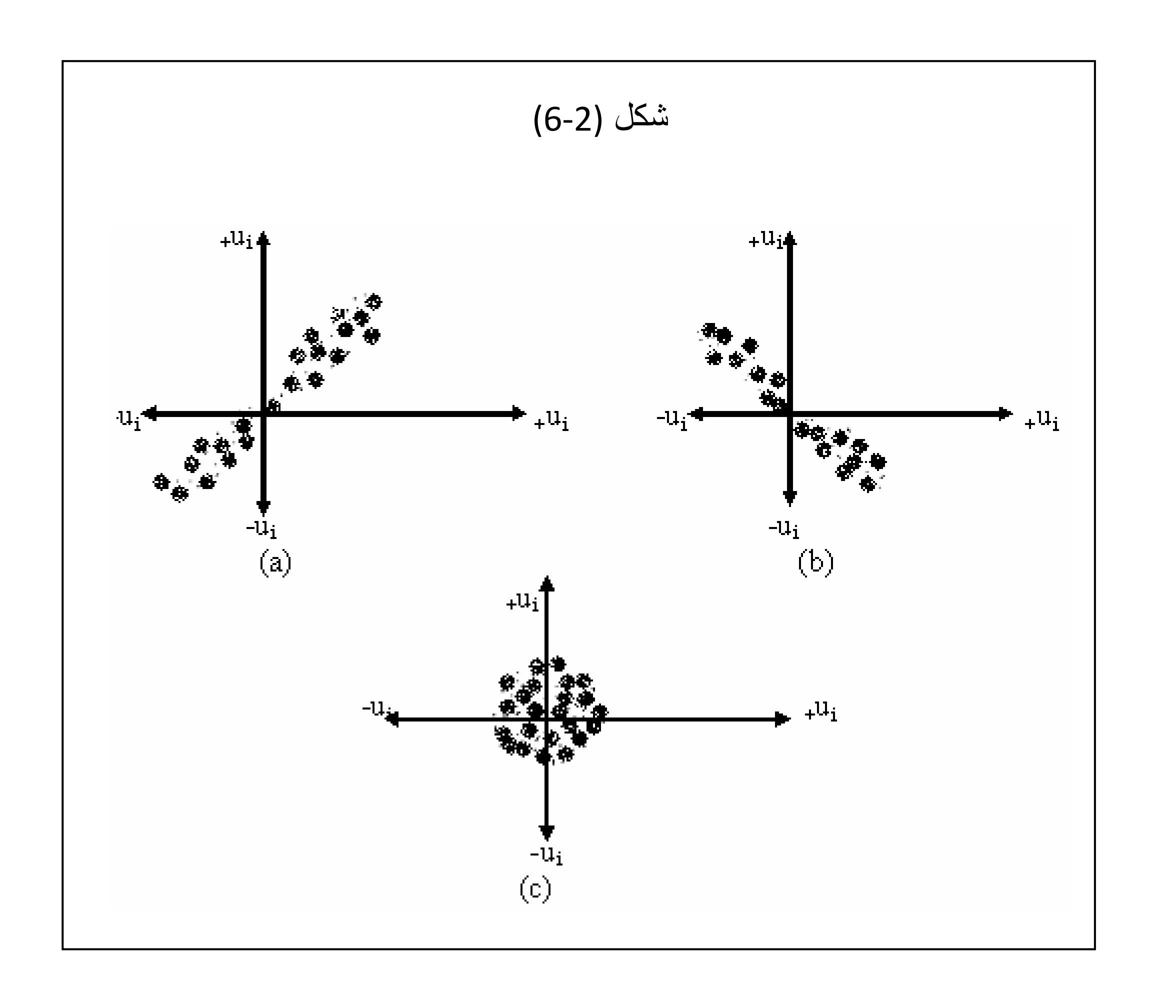
لايوجد ارتباط ذاتي بين مشاهدات المتغير العشوائي.

أي:

$$\begin{aligned}\text{cov}(u_i, u_j / X_i, X_j) &= \text{E}(u_i - \text{E}(u_i)/X_i)(u_j - \text{E}(u_j)/X_j) \\ &= \text{E}(u_i/X_i)(u_j/X_j) \\ &= 0\end{aligned}$$

j, i مشاهدتان مختلفتان.

وهذه الفرضية أيضاً تسمى لا وجود للارتباط المتسلسل no serial correlation ويمكن توضيح الأنماط المختلفة فيما بين المتغير العشوائي من الأشكال التالية:



شكل (2-6)

في الشكل (2-6)،(a) يتضح ان قيم u مرتبطة ذاتياً باتجاه طردي متزايد. حيث ان القيم الموجبة من u تتبعها قيم موجبة من u. في حين القيم السالبة تتبع بقيم سالبة أيضاً.

في الشكل (b) فان قيم u مرتبطة ذاتياً بشكل عكسي، اذ ان قيم u الموجبة تتبعها قيم u السالبة والعكس صحيح.

وبذلك فالشكلان (a) و (b) تعكس الارتباط الذاتي او المتسلسل.

في حين الشكل (c) يوضح عدم وجود الارتباط الذاتي اذ لايتوضح وجود نمط من نوع معين.

<u>الفرض (٦) :</u>

لا وجود للارتباط بين u_i و X_i .

أو
$$\text{cov}(u_i, X_i) = \text{E}(u_i, X_i) = 0$$

وذلك من أجل عزل آثار المتغير X والمتغير u على المتغير المعتمد Y .

ويتحقق هذا الافتراض في حالتين:

(١) اذا كان المتغير X غير عشوائي.

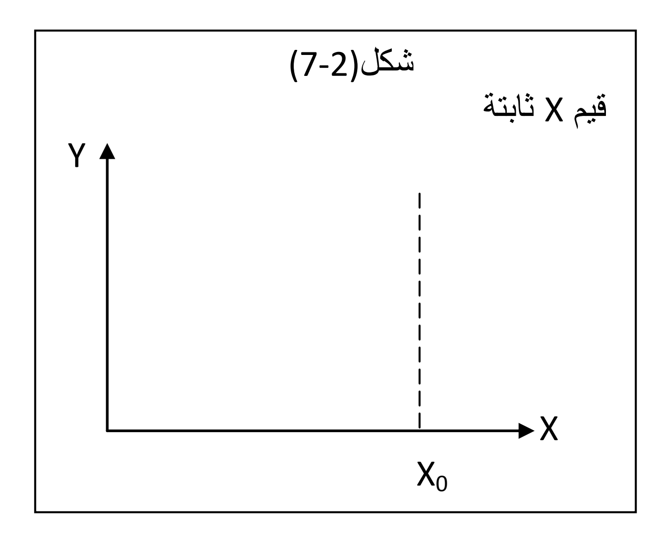
(٢) اذا كان المتغير X عشوائي لكنه غير مرتبط مع u .

الفرض (٧) :

عدد المشاهدات (n) يجب ان تكون اكبر من عدد المعلمات المطلوب تقديرها.

الفرض (٨) :

ان المتغيرات المستقلة (X) تقاس بدون أخطاء، كما ان قيم X المستخدمة في العينة تتطلب ان تمتلك تغيرات ملموسة فإذا لا توجد تغيرات ملموسة في (X)، لا يكون بالإمكان من توضيح التغيرات في المتغير المعتمد (Y) فإذا كانت قيم X ثابتة عند X_0 كما في الشكل (2-7) فان المربعات الصغرى غير صالحة للاستخدام.



الفرض (٩) :

علاقة الانحدار يجب ان تكون موصفة بشكل صحيح، ولاتوجد أخطاء للتوصبف من ناحية المتغيرات التوضيحية او من ناحية الشكل الدالي المستخدم (خطي بدلالة المعلمات او المتغيرات او كلاهما) او من حيث عدد المعادلات المستخدمة (منفردة أو آنية).

الفرض (١٠) :

خلو النموذج من أخطاء التجميع للمتغيرات وأخطاء القياس للبيانات.

الفرض (١١) :

ان المتغير العشوائي له توزيع طبيعي Normal. أي يكون التوزيع عند كل نقطة من X متماثلاً حول الوسط الحسابي. حيث ان التوزيع الطبيعي هو أقرب للواقع كما ان افتراضه يسهل استخدام اختبارات اخرى وكما سيتم توضيحه في الفقرات اللاحقة.

(3-4-2)خواص المقدرات بطريقة المربعات الصغرى (Proporties of least squares estimates)

نظرية جاوس – ماركوف

حيث ان معلمة الانحدار تمثل:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\Sigma xy}{\Sigma x^2}$$

$$= \frac{\Sigma x(Y - \bar{Y})}{\Sigma x^2} = \frac{\Sigma xY - Y\Sigma x}{\Sigma x^2}$$

حيث $(\Sigma x_i = 0)$

$$= \frac{\Sigma xY}{\Sigma x^2}$$

$$= \sum_{i=1}^{n} \frac{x_i}{\Sigma x_i^2} \cdot Y_i$$

$$= \sum_{i=1}^{n} k_i \cdot Y_i$$

حيث $k_i = \frac{x_i}{\Sigma x_i^2}$

وبذلك فان $\hat{\beta}_1$ مقدر خطي لانه دالة خطية بدلالة Y حيث انها تركيب خطي بدلالة Y وان اوزان التركيب هي k_i .

وبالمثل فان $\hat{\beta}_0$ هي تركيب خطي أيضاً بدلالة Y . (يترك البرهان واجب)

فالخاصية الأولى: ان المعلمات المقدرة هي تراكيب خطية بدلالة Y .

واجب بيتي: اثبت ان $\hat{\beta}_0$ دالة خطية بدلالة مشاهدات Y.

<u>الخاصية الثانية:</u> عدم التحيز (unbiased ness)

أي ان:

$$\mathrm{E}(\hat{\beta}_i) = \beta_i \qquad \forall \qquad i = 0,1$$

وبالنسبة للمعلمة $\hat{\beta}_1$:

$$\hat{\beta}_1 = \Sigma k_i Y_i$$

$$\hat{\beta}_1 = \Sigma k_i (\beta_0 + \beta_1 X_i + u_i)$$

$$\hat{\beta}_1 = \beta_0 \Sigma k_i + \beta_1 k_i X_i + \Sigma k_i u_i$$

$$\hat{\beta}_1 = \beta_1 + \Sigma k_i u_i \qquad \cdots \qquad (13-2)$$

حيث ان :

$$\Sigma k_i = \Sigma \frac{x_i}{\Sigma x_i^2} = \frac{1}{\Sigma x_i^2} \cdot \Sigma x_i$$

لان Σx_i^2 قيمة معلومة

و $(\Sigma x_i = 0)$، مجموع الانحرافات عن المتوسط

$$\therefore \Sigma k_i = 0$$

كما ان :

$$\Sigma k_i X_i = \Sigma \frac{x_i}{\Sigma x_i^2} X_i$$

$$= \Sigma \frac{x_i}{\Sigma x_i^2} (x_i + \overline{X})$$

$$= \Sigma \frac{x_i^2 + x_i \overline{X}}{\Sigma x_i^2}$$

$$= \frac{\Sigma x_i^2}{\Sigma x_i^2} + \overline{X} \frac{\Sigma x_i}{\Sigma x_i^2} \qquad ; \qquad \Sigma x_i = 0$$

$$= 1$$

وبأخذ التوقع لطرفي المعادلة (13-2) :

$$\mathrm{E}(\hat{\beta}_1) = \beta_1 + \sum_i k_i \mathrm{E}(u_i)$$

$$= \beta_1$$

حيث $\mathrm{E}(u_i) = 0$ بموجب فروض التحليل

وبذلك فان $\hat{\beta}_1$ معلمة غير متحيزة.

وبالمثل يمكن البرهنة على ان $\hat{\beta}_0$ غير متحيزة ايضاً أي:

$$\mathrm{E}(\hat{\beta}_0) = \beta_0$$

(تمرين للطالب)

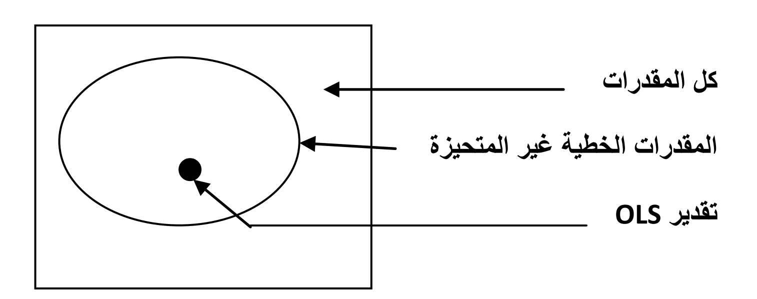
<u>اما الخاصية الثالثة :</u> تباين المعلمات المقدرة هو أقل ما يمكن من بين مجموع المقدرات غير المتحيزة والخطية، (يترك البرهان).

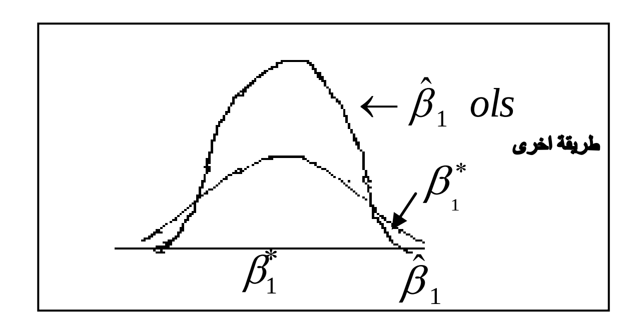
ولذلك تسمى المقدرات بموجب المربعات الصغرى الاعتيادية (BLUE) تشمل الصفات الثلاث ، خطية وغير متحيزة ولها أقل تباين. وهذه ما يطلق عليها نظرية جاوس– ماركوف والتي تنص على الآتي: (مع

تحقق فرضيات النموذج الخطي فان المقدرات بموجب طريقة المربعات الصغرى تمثل أقل تباين ضمن مجموعة المقدرات الخطية وغير المتحيزة).

ويمكن توضيح ما تنص عليه هذه النظرية من الأشكال التالية:

شكل فن Vein diagram





(4-4-2) طريقة الإمكان الأعظم. " Maximum Likelihood "

ويرمز لهذه الطريقة اختصاراً (ML) . اما المعيار الذي تستند اليه هذه الطريقة فهو تعظيم دالة الامكان (Likelihood function) .

ودالة الامكان ببساطة تمثل دالة الكثافة الاحتمالية المشتركة للمتغير العشوائي Y ولذلك نحتاج لمعرفة توزيع ذلك المتغير Y.

وبموجب الافتراضات (1) و (2) فان توزيع Y هو توزيع u نفسه.

وبموجب الفرض (11) فان توزيع Y هو توزيع طبيعي.

ولتحديد متوسط Y وتباينه:

$$\mathrm{E}(Y) = \beta_0 + \beta_1 X \quad \& \quad \mathrm{var}(Y) = \sigma^2$$

اذن $Y_i \sim N(\beta_0 + \beta_1 X_i \ ; \ \sigma^2)$

ولذلك فان دالة الكثافة الاحتمالية المشتركة لمشاهدات Y المشروطة لـ $\beta_0 + \beta_1 X_i$ و σ^2

ويرمز لها: $f(Y_1, Y_2 \ . \ . \ . \ , Y_n / \beta_0 + \beta_1 X_i \ , \ \sigma^2)$.

وحيث ان مشاهدات Y مستقلة الواحدة عن الأخرى، لذلك فان الكثافة الاحتمالية المشتركة تعد حاصل ضرب الكثافات الحدية المنفردة لمشاهدات Y :

$$f(Y_1, Y_2 \; . \; . \; . \; , Y_n / \beta_0 + \beta_1 X_i \;,\; \sigma^2) = f(Y_1 / \beta_0 + \beta_1 X_i \;,\; \sigma^2) \cdot f(Y_2 / \beta_0 + \beta_1 X_i \;,\; \sigma^2)$$

$$. \quad . \quad . \quad \cdot f(Y_n / \beta_0 + \beta_1 X_i \;,\; \sigma^2).$$

وبافتراض التوزيع الطبيعي، فان

$$f(Y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \;\; \exp \;\left\{ -\frac{1}{2} \frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{\sigma^2} \right\}$$

اي:

$$f(Y_1, Y_2 \; . \; . \; . \; , Y_n / \beta_0 + \beta_1 X_i \;,\; \sigma^2) = \underset{i=1}{\overset{n}{\pi}} f(Y_i)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \;\left\{ -\frac{1}{2} \frac{\Sigma (Y_i - \beta_0 - \beta_1 X_i)^2}{\sigma^2} \right\}$$

أي ان دالة الإمكان:

$$LF(\beta_0, \beta_1, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \;\left\{ -\frac{1}{2} \frac{\Sigma (Y_i - \beta_0 - \beta_1 X_i)^2}{\sigma^2} \right\} \;\; . \;\; . \;\; .$$

وللتبسيط نأخذ لوغاريتم الطرفين:

$$\ln LF = \frac{-n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) \;\; - \frac{1}{2} \Sigma \frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{\sigma^2} \qquad . \; . \; . \qquad (14\text{-}2)$$

وبتطبيق شروط الامثلية:

الشرط الضروري:

$$\frac{\partial \ln LF}{\partial \beta_0} = -\frac{1}{\sigma^2} \Sigma (Y_i - \beta_0 - \beta_1 X_i)(-1) = 0 \qquad . \; . \; . \qquad (15\text{-}2)$$

$$\frac{\partial \ln LF}{\partial \beta_1} = -\frac{1}{\sigma^2} \Sigma (Y_i - \beta_0 - \beta_1 X_i)(-X) = 0 \qquad . \; . \; . \qquad (16\text{-}2)$$

$$\frac{\partial \ln LF}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \Sigma (Y_i - \beta_0 - \beta_1 X_i)^2 = 0 \qquad . \; . \; . \qquad (17\text{-}2)$$

وبعد تبسيطها. فان المعادلتين (2-15) و (2-16) هي المعادلات الطبيعية نفسها التي تم الحصول عليها عند استخدام طريقة المربعات الصغرى الاعتيادية. وبذلك فان قيم $\hat{\beta}_1, \hat{\beta}_0$ لها القوانين نفسها المستخدمة بطريقة المربعات الصغرى الاعتيادية وذلك ينسجم مع كون الحد الأخير في المعادلة (2-14) مسبوقاً

بإشارة سالبة وعليه فان تعظيم دالة الإمكان يمثل تصغيراً لهذا الحد، والذي ينطبق مع المربعات الصغرى.
وبعد نعويض قيم $\hat{\beta}_1, \hat{\beta}_0$ في المعادلة (2-17) وتبسيطها:

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\Sigma e_i^2 \qquad ... \qquad (18-2)$$

وبأخذ التوقع لطرفي المعادلة (2-18) :

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \frac{1}{n}\mathrm{E}(\Sigma e_i^2)$$

$$= \frac{n-2}{n}\sigma^2 = \sigma^2 - \frac{2}{n}\sigma^2 \qquad ... \qquad (19-2)$$

ويتبين من المعادلة (2-19) ان تقدير تباين الخطأ بموجب طريقة الإمكان الأعظم متحيز نحو الأسفل. بمعنى ان الإمكان الأعظم يعطي مقدراً لتباين الخطأ أقل من قيمته الحقيقية σ^2 وخاصة في العينات الصغيرة. في حين مع كبر حجم العينة فان مقدار التحيز سيضمحل ويقترب من قيمته الحقيقية.

(2-4-5) <u>تباين المعلمات المقدرة</u>: <u>Variance of the estimated parameters</u>

توضح المعادلات الطبيعية بان المقدرات $\hat{\beta}_1, \hat{\beta}_0$ دوال بدلالة بيانات العينة التي يتم استخدامها، غير ان البيانات تتغير من عينة إلى أخرى، لذا فان التقديرات ستتغير أيضاً ، ولذلك يتطلب مقياساً لدقة المعلمات المقدرة. وإحصائياً فان مقياس الدقة يعبر عنه بواسطة الانحراف المعياري.
واعتماداً على فرضيات جاوس فان تباين المعلمة المقدرة β_1 على وفق الآتي:

$$\mathbf{var}(\hat{\beta}_1) = \mathbf{E}(\hat{\beta}_1 - \mathbf{E}(\hat{\beta}_1))^2$$

وحيث ان $\hat{\beta}_1$ غير متحيزة فان:

$$\mathbf{var}(\hat{\beta}_1) = \mathbf{E}(\hat{\beta}_1 - \beta_1)^2$$

وبالاعتماد على العلاقة (2-13)

$$\mathrm{var}(\hat{\beta}_1) = \mathrm{E}(\Sigma k_i u_i)^2$$

$$\mathrm{var}(\hat{\beta}_1) = \mathrm{E}(k_1^2 u_1^2 + k_2^2 u_2^2 + ... + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + ... + 2k_{n-1} k_n u_{n-1} u_n)$$

$$\mathrm{var}(\hat{\beta}_1) = \sigma^2 \Sigma k_i^2$$

وذلك بالاعتماد على فروض التحليل:

$E(u_i^2) = \sigma^2$ لجميع قيم i (الفرض 4)

$E(u_i u_j) = 0$, $i \neq j$ (الفرض 5)

وبذلك:

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma x_i^2} = \frac{\sigma^2}{S_{xx}} \qquad . . . \qquad (20-2)$$

وذلك لان:

$$\Sigma(k_i^2) = \Sigma(\frac{x_i}{\Sigma x_i^2})^2$$

$$= \Sigma \frac{x_i^2}{(\Sigma x_i^2)^2} = \frac{1}{\Sigma x_i^2}$$

ويتضح من العلاقة (2-20) ان تباين $\hat{\beta}_1$ يتناسب طردياً مع (σ^2) وعكسياً مع Σx_i^2 . فكلما كانت التغيرات في المتغير المستقل كبيرة فان التباين يكون صغيراًَ وبالتالي تزداد دقة المعلمة β_1 المقدرة. اما اذا كان تباين الخطأ كبيراً فان دقة المعلمة المقدرة ستنخفض كما ان زيادة حجم العينة (n) يؤدي بدوره الى ان عدد الحدود في Σx_i^2 يتزايد وبذلك يكون التباين منخفضاً بمعنى تتحسن دقة المعلمة المقدرة.

ومن جانب آخر فان الجذر التربيعي للتباين يعبر عن الانحراف المعياري للمعلمة:

$$s.e(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\Sigma x_i^2}}$$

وبالمثل يمكن اشتقاق تباين معلمة المقطع الصادي المقدرة $\hat{\beta}_0$ وانحرافها المعياري:

$$\text{var}(\hat{\beta}_0) = \frac{\Sigma X_i^2}{n \Sigma x_i^2} \cdot \sigma^2 \qquad . . . \qquad (21-2)$$

$$= \sigma^2 (\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x_i^2})$$

(البرهان يترك للطالب)

العلاقة (2-21) تؤكد على ان تباين معلمة الثابت المقدرة $\hat{\beta}_0$ تتناسب طردياً σ^2 و ΣX_i^2 وعكسياً مع Σx_i^2 وحجم العينة n .

كما ان الانحراف المعياري لهذه المعلمة:

$$s.e(\hat{\beta}_0) = \sqrt{\frac{\Sigma X_i^2}{n\Sigma x_i^2}} \cdot \sigma$$

(6-4-2)التباين المشترك بين $\hat{\beta}_1, \hat{\beta}_0$: Covariance between $\hat{\beta}_1, \hat{\beta}_0$

معلوم ان المقدرات $\hat{\beta}_1, \hat{\beta}_0$ لاتتغير من عينة الى أخرى وانما أيضاً في العينة المعطاة في الغالب ترتبط هذه المعلمات مع بعضها. وان هذا الارتباط يمكن ان يقاس من خلال التباين المشترك بينهما. واعتماداً على تعريف التباين المشترك:

$$\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{E}\left\{ \left[\hat{\beta}_0 - \text{E}(\hat{\beta}_0)\right] \left[\hat{\beta}_1 - \text{E}(\hat{\beta}_1)\right] \right\}$$

وبما ان المعلمات غير متحيزة:

$$= \text{E}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) \qquad ... \qquad (22\text{-}2)$$

ولكن

$$\hat{\beta}_0 - \text{E}(\hat{\beta}_0) = \hat{\beta}_0 - \text{E}(\bar{y} - \hat{\beta}_1\bar{X})$$

$$= \bar{y} - \hat{\beta}_1\bar{X} - \bar{y} + \beta_1\bar{X}$$

$$= -\bar{X}(\hat{\beta}_1 - \beta_1)$$

تعوض في العلاقة (2-22):

$$\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X} \ \text{E}(\hat{\beta}_1 - \beta_1)^2$$

$$= -\bar{X} \ \text{var}(\hat{\beta}_1)$$

$$= -\bar{X} \ \frac{\sigma^2}{\Sigma x_i^2} \qquad ... \qquad (23\text{-}2)$$

وحيث ان $\text{var}(\hat{\beta}_1)$ دائماً موجب (صفة أي متغير) فان طبيعة الارتباط بين $\hat{\beta}_1, \hat{\beta}_0$ يعتمد على اشارة متوسط $(\bar{X})$ X. فاذا كان $\bar{X}$ موجباً فان التباين المشترك سيكون سالباً.

يتضح من العلاقات (2-20) ، (2-21) ، (2-23) ان التباين والتباين المشترك للمعلمات المقدرة يعتمد على σ^2 وهو تباين المتغير العشوائي للمجتمع وهو بدوره غير معلوم لذا لابد من ايجاد طريقة لتقديره.

(7-4-2)تقدير تباين الخطأ للمجتمع $(\hat{\sigma}^2)$ بموجب المربعات الصغرى. The estimate of population variance error

بالرجوع للمعادلة (2-3) :

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

وبعد أخذ المتوسط لطرفي المعادلة:

$$\bar{Y} = \beta_0 + \beta_1\bar{X} + \bar{u}$$

وبالطرح ينتج:
$$Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (u_i - \bar{u})$$
$$y_i = \beta_1 x_i + (u_i - \bar{u})$$
وحيث ان $\hat{u}_i = y_i - \hat{\beta}_1 x_i$ ، $\Rightarrow \hat{u}_i = \beta_1 x_i + (u_i - \bar{u}) - \hat{\beta}_1 x_i$
$$= -(\hat{\beta}_1 - \beta_1) x_i + (u_i - \bar{u})$$
وبتربيع الطرفين وأخذ التجميع:
$$\Sigma \hat{u}_i^2 = (\hat{\beta}_1 - \beta_1)^2 \Sigma x_i^2 + \Sigma (u_i - \bar{u})^2 - 2(\hat{\beta}_1 - \beta_1) \Sigma x_i (u_i - \bar{u})$$
وبأخذ التوقع:
$$E(\Sigma \hat{u}_i^2) = \Sigma x_i^2 E(\hat{\beta}_1 - \beta_1)^2 + E(\Sigma (u_i - \bar{u})^2) - 2E(\hat{\beta}_1 - \beta_1) \Sigma x_i (u_i - \bar{u}) \quad . . . \quad (24\text{-}2)$$
$$= \Sigma x_i^2 \operatorname{var}(\hat{\beta}_1) + (n-1)\sigma^2 - 2\sigma^2$$
$$= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$
$$\hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{n-2} = \frac{\Sigma e^2}{n-2} \quad . . . \quad (25\text{-}2)$$
وهي غير متحيزة أيضاً لان $E(\hat{\sigma}^2) = \sigma^2$

الحد الثاني من العلاقة (24-2) يمكن تبسيطه كالآتي:
$$E\left[\Sigma (u_i - \bar{u})^2\right] = E\left[\Sigma (u_i^2 - 2u_i \bar{u} + \bar{u}^2)\right]$$
$$= E\left[\Sigma u_i^2 - 2\bar{u} \Sigma u_i + \Sigma \bar{u}^2\right]$$
$$= E\left[\Sigma u_i^2 - 2\frac{\Sigma u_i}{n} \Sigma u_i + n \Sigma \bar{u}^2\right]$$
$$= E\left[\Sigma u_i^2 - 2\frac{(\Sigma u_i)^2}{n} + n\left(\frac{\Sigma u_i}{n}\right)^2\right]$$
$$= E\left[\Sigma u_i^2 - 2\frac{(\Sigma u_i)^2}{n} + \frac{(\Sigma u_i)^2}{n}\right]$$
$$= E\left[\Sigma u_i^2 - \frac{(\Sigma u_i)^2}{n}\right]$$
$$= \Sigma E(u_i^2) - E\frac{(\Sigma u_i)^2}{n}$$
$$= n\sigma^2 - n\, E\, \bar{u}^2$$

وحيث ان:

$$E(\bar{u}^2) = \frac{\sigma^2}{n}$$

$$E\Sigma(u_i - \bar{u})^2 = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

اما الحد الثالث من العلاقة (2-24) فيمكن تبسيطه كالآتي :

$$-2E(\hat{\beta}_1 - \beta_1)\Sigma x_i(u_i - \bar{u}) = -2E\left[(\hat{\beta}_1 - \beta_1)\Sigma x_i u_i - \bar{u}(\hat{\beta}_1 - \beta_1)\Sigma x_i\right]$$

$$= -2E\left[(\hat{\beta}_1 - \beta_1)\Sigma x_i u_i\right] \quad ... \quad (26-2)$$

حيث ان :

$$\hat{\beta}_1 = \frac{\Sigma x_i y_i}{\Sigma x_i^2} = \frac{\Sigma x_i(\beta_0 + \beta_1 X_i + u_i)}{\Sigma x_i^2}$$

$$\hat{\beta}_1 \Sigma x_i^2 = \beta_0 \Sigma x_i + \beta_1 \Sigma x_i X_i + \Sigma x_i u_i$$

$$= \beta_1 \Sigma x_i(x_i + \bar{X}) + \Sigma x_i u_i$$

$$= \beta_1 \Sigma x_i^2 + \beta_1 \bar{X}\Sigma x_i + \Sigma x_i u_i$$

$$\therefore \hat{\beta}_1 \Sigma x_i^2 - \beta_1 \Sigma x_i^2 = \Sigma x_i u_i$$

$$(\hat{\beta}_1 - \beta_1)\Sigma x_i^2 = \Sigma x_i u_i$$

بالتعويض عن $(\Sigma x_i u_i)$في العلاقة (2-26) نحصل على:

$$-2E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta)\Sigma x_i^2 = -2\frac{\sigma^2}{\Sigma x_i^2}\Sigma x_i^2$$

$$= -2\sigma^2$$

وبالعودة إلى خطوات بناء نموذج انحدار خطي بسيط في المبحث (2-3) فبعد تقدير معلمات النموذج ننتقل إلى خطوة الاستدلال حول صحة النتائج التقديرية ونحتاج لذلك حساب الانحراف المعياري للمعلمات وكذلك الانحراف المعياري للخطأ.

وباستعادة المعلومات حول تباين المعلمات:

$$\text{var}(\hat{\beta}_0) = \hat{\sigma}^2 \cdot \frac{\Sigma X_i^2}{nS_{xx}} \quad \& \quad \text{var}(\hat{\beta}_1) = \hat{\sigma} \cdot \frac{1}{S_{xx}}$$

علماً ان:

$$\hat{\sigma}^2 = MSE = \frac{\Sigma e^2}{n-2}$$

وهذا يتطلب حساب الآتي:

(١) $\hat{Y}_i$ ، $\hat{Y}_i = 35.35 + 2.564 X_i$ لكل مشاهدة من مشاهدات الجدول (2-١) .

(٢) ثم تحسب البواقي $e_i = Y_i - \hat{Y}_i$. كما موضحة في الجدول (2-3)

جدول (2-3)

المزرعة	$\hat{Y}$	$e = Y - \hat{Y}$
1	163.55	-23.55
2	548.15	-48.15
3	317.39	82.61
4	240.47	59.53
5	343.03	12.97
6	226.368	14.132
7	263.2896	-62.6896
8	49.9648	-16.4648
9	63.554	6.246
10	43.5548	-24.8548

(٣) تحسب مجموع مربعات البواقي $\Sigma e_i^2 = 18467.04$

وبذلك فان تباين الخطأ المقدر هو:

$$\hat{\sigma}^2 = \frac{\Sigma e^2}{n-2} = 2308.38$$

ولتقدير تباين المعلمات المقدرة:

$$\Sigma X_i^2 = 89017.19$$

$$\Sigma x_i^2 = \Sigma X_i^2 - n\bar{X}^2$$

$$= 89017.19 - 10(74.33)^2 = 89017.19 - 55249.489$$

$$= 33767.701$$

$$\Rightarrow S_{xx} = 33767.701$$

$$\text{var}(\hat{\beta}_0) = 2308.38 \cdot \frac{89017.19}{10(33767.701)} = 608.53$$

وهكذا فان الانحراف المعياري للمعلمة $(\hat{\beta}_0)$:

$$s.e(\hat{\beta}_0) = 24.668$$

وكذلك بالنسبة للمعلمة $(\hat{\beta}_1)$ فان:

$$\text{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}} = \frac{2308.38}{33767.701} = 0.068$$

وبذلك فان انحرافها المعياري:

$$s.e(\hat{\beta}_1) = 0.261$$

(8-4-2) <u>طريقة أخرى لحساب مجموع مربعات الخطأ</u> (Σe_i^2) <u>Calculation of sum of squared error:</u>

$$e_i = Y_i - \hat{Y}_i$$

$$= Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

$$= Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) - \hat{\beta}_1 X_i$$

$$= Y_i - \bar{Y} - \hat{\beta}_1 (X_i - \bar{X})$$

$$= y_i - \hat{\beta}_1 x_i$$

بتربيع الطرفين وأخذ المجموع:

$$\Sigma e_i^2 = \Sigma y_i^2 - 2\hat{\beta}_1 \Sigma x_i y_i + \hat{\beta}_1^2 \Sigma x_i^2$$

$$= \Sigma y_i^2 - 2\hat{\beta}_1 \Sigma x_i y_i + \hat{\beta}_1 \frac{\Sigma x_i y_i}{\Sigma x_i^2} \Sigma x_i^2$$

$$= \Sigma y_i^2 - \hat{\beta}_1 \Sigma x_i y_i \qquad \cdots \qquad (27-2)$$

$$\Sigma e_i^2 = \Sigma y_i^2 - 2\hat{\beta}_1 (\hat{\beta}_1 \Sigma x_i^2) + \hat{\beta}_1 \Sigma x_i^2$$

$$= \Sigma y_i^2 - \hat{\beta}_1^2 \Sigma x_i^2 \qquad . . . \qquad (28-2)$$

أو:

$$= \Sigma y_i^2 - \frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2} \qquad . . . \qquad (29-2)$$

(Σy_i^2) : تمثل مجموع مربعات التغيرات في مشاهدات المتغير y .

ويرمز لها (TSS) تمثل مجموع التغيرات الاجمالية في مشاهدات المتغير y مقاسة حول المتوسط.

أما: $\hat{\beta}_1^2 \Sigma x_i^2$ أو $\hat{\beta}_1 \Sigma x_i y_i$ أو $\frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}$ تسمى التغيرات المشروحة من قبل معادلة الانحدار ويرمز لها (ESS) .

(Σe_i^2) : تشير الى مجموع مربعات البواقي ويرمز لها (RSS) وتمثل التغيرات غير المشروحة.

بعبارة أخرى: $TSS = ESS + RSS$

ان مجموع التغيرات الاجمالية تم تقسيمه إلى مركبتين: التغيرات المشروحة + التغيرات غير المشروحة.

وبالرجوع الى بيانات الجدول(2-١) واستخدام الصيغ الرياضية يتم حساب الآتي:

$$(TSS) = \Sigma y_i^2 = \Sigma(Y_i - \overline{Y})^2 = \Sigma Y_i^2 - n\overline{Y}^2$$

$$= 750760.6 - 10(51035.3281)$$

$$= 750760.6 - 510353.281$$

$$= 240407.309$$

$$ESS = \Sigma \hat{y}_i^2 = \hat{\beta}_1 \Sigma x_i y_i = \hat{\beta}_1 \left[\Sigma X_i Y_i - n\overline{X}\overline{Y}\right]$$

$$= 2.564(254489.2 - 10(74.33)(225.91)$$

$$= 221966.242$$

$$\Rightarrow \Sigma e_i^2 = TSS - ESS = 18441.067$$

مناقشة: بالاعتماد على بيانات الجدول (2-4) :

ناقش الآتي:

جدول (2-4)

X	١	٢	٣	٤	٥	٦
Y	٤	٦	٧	٧	٩	١١

(١) ارسم شكل الانتشار؟

(٢) استخدم القوانين لتقدير الميل والمقطع الصادي للخط المقدر بطريقة OLS وارسم الخط المقدر؟

(٣) احصل على متوسط Y ومتوسط X واحصل على قيمة Y عندما $X = \overline{X}$ وحددها على الرسم، مالذي تستنتجه من هذه القيمة المقدرة؟

(٤) استخدم التقديرات التي حصلت عليها في (٢) واحسب بواقي التقدير ثم احسب مجموع مربعات البواقي ؟

(٥) ما شكل الانحدار اذا β_0 = صفر (أ) بالرسم ، (ب) جبرياً

(٦) هل ان قوانين تقدير المعلمات بموجب OLS المعادلات (2-10) و (2-11) تصبح ملائمة اذا علمت ان $\beta_0 = 0$ ؟

(٧) اذا $\beta_0 = 0$ حدد مجموع مربعات البواقي جبرياً، ثم استخدم العينة لتقدير β_1 ؟

(٨) استخدم التقدير الذي حصلت عليه في (٧) وارسم الخط المقدر على ورقة الرسم نفسها للانتشار. ما استنتاجك؟

(٩) استخدم التقدير في (٧) واحصل على البواقي ثم احسب مجموع مربعات البواقي؟

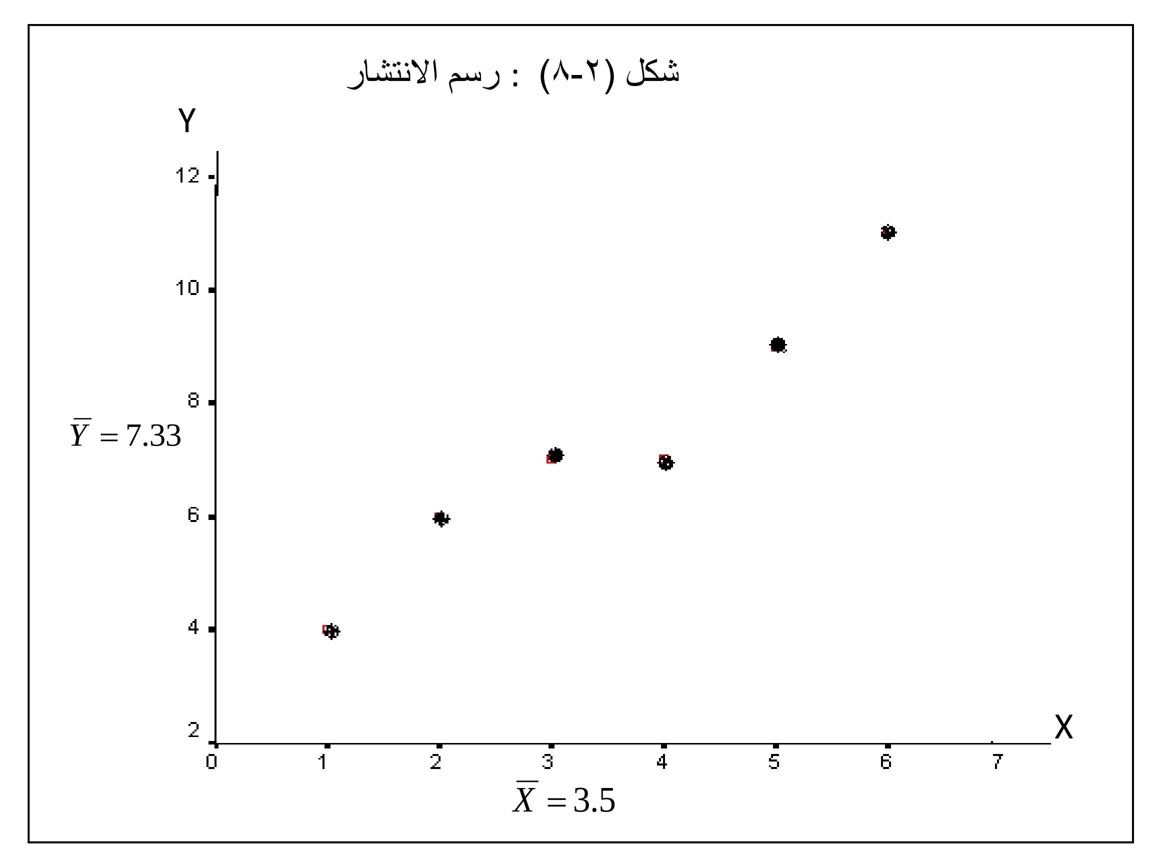
(١٠) قارن بين النتائج في (٤) و (٩) ؟

(١١) افترض ان القيمة الحقيقية لـ β_1 = صفر ما شكل الانحدار (أ) جبرياً ، (ب) بالرسم

(١٢) ما نوع العلاقة بين Y و X اذا $\beta_1 = 0$ ؟

الحل:

(١) رسم الانتشار



رسم الانتشار لبيانات جدول (2-4) باستخدام برنامج SPSS

(٢) قوانين تقدير الميل والمقطع الصادي بواسطة (OLS) هي:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad ، \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$ حيث ان:

$$S_{xy} = \Sigma(y - \bar{y})(X - \bar{X}) = \Sigma xy$$

$$S_{xx} = \Sigma(X - \bar{X})^2 = \Sigma x^2$$

ولحسابها يتطلب جدول الحسابات (2-5).

جدول (2-5)
جدول الحسابات بالقيم كانحرافات عن المتوسطات

رقم المشاهدة	Y	X	$y = Y - \bar{Y}$	$x = X - \bar{X}$	xy	x^2
1	٤	١	-3.33	-2.5	8.325	6.25
2	6	٢	-1.33	-1.5	`1.995	2.25
3	٧	٣	-.33	-.5	0.165	0.25
4	٧	٤	-.33	.5	0.165	0.25
5	٩	٥	1.67	1.5	2.505	2.25
6	١١	٦	3.67	2.5	9.175	6.25
Σ	٤٤	٢١	0.02	0.0	22.33	17.5

$\bar{Y} = 7.33$

$\bar{X} = 3.5$

$S_{xy} = 22.33$ ، $S_{xx} = 17.5$ ، $\hat{\beta}_1 = 1.276 \approx 1.28$

$\hat{\beta}_0 = 7.33 - 1.28(3.5) = 2.85$

معادلة التقدير:

$$\hat{Y}_i = 2.85 + 1.28 X_i$$

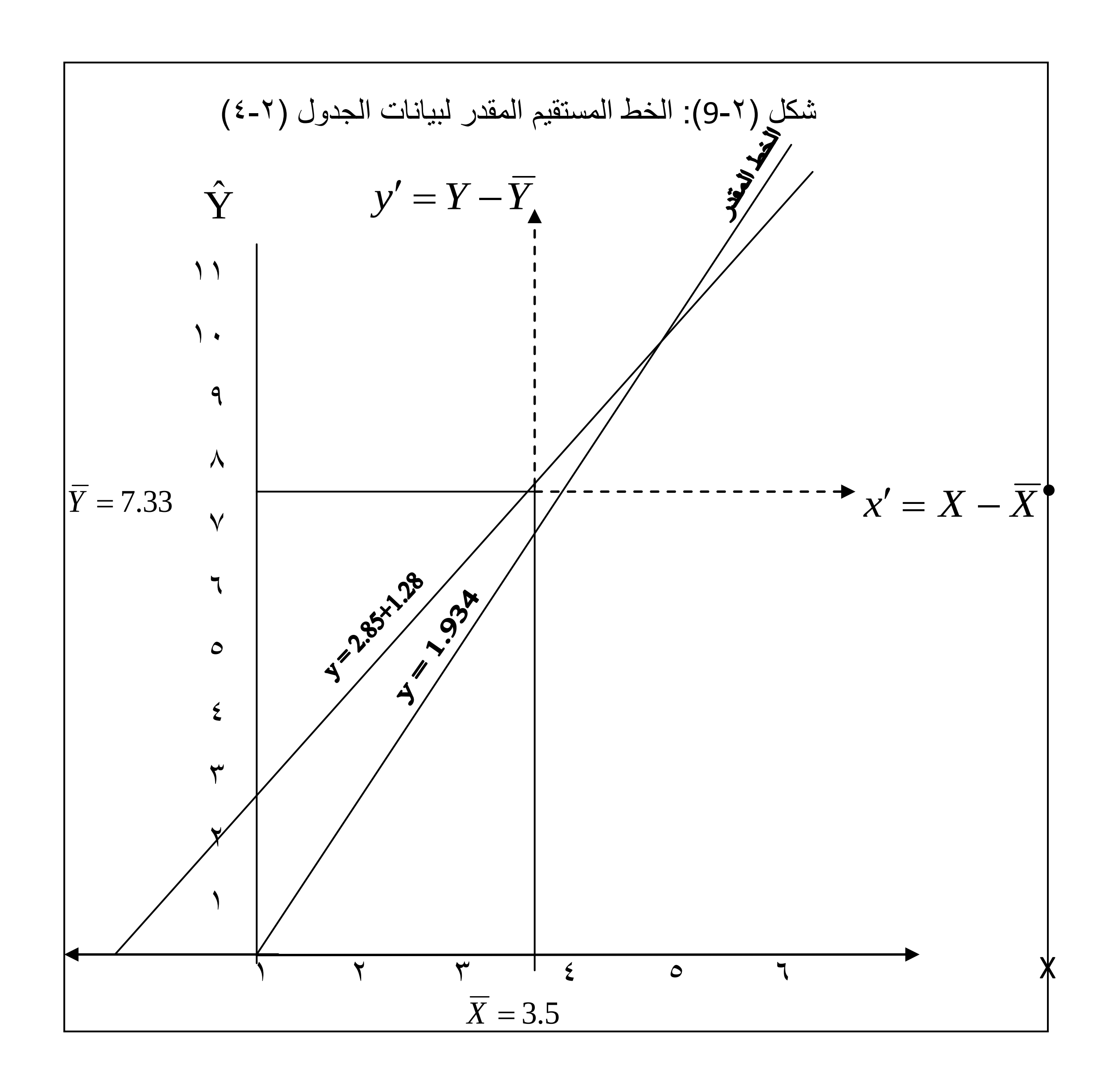
لرسم الخط المستقيم:

نفرض:

$$X = 0 \quad \Rightarrow \quad \hat{Y}_i = 2.85 \quad ،$$

$$\hat{Y} = 0 \quad \Rightarrow \quad X = -2.23$$

وبتعيين النقطتين على الشكل ونوصل بينهما نحصل على الخط المستقيم المقدر:



شكل (٢-9): الخط المستقيم المقدر لبيانات الجدول (٢-٤)

(٣) $\bar{Y} = 7.33$ ، $\bar{X} = 3.5$

$$(\hat{Y}_i / X = 3.5) = 2.85 + 1.28(3.5) = 7.33$$

نستنتج ان نقطة المتوسط $(\bar{X}, \bar{Y})$ تقع على الخط المستقيم المقدر.

(٤) اعتماداً على معادلة التقدير $(\hat{Y}_i = 2.85 + 1.28X_i)$ يتم حساب قيم $(\hat{Y}_i)$ لكل قيمة من قيم X_i كالآتي:

$$\hat{Y}_1 = 2.85 + 1.28(1) = 4.13$$

$$\hat{Y}_2 = 2.85 + 1.28(2) = 5.41$$

$$\vdots$$

$$\hat{Y}_6 = 2.85 + 1.28(6) = 10.35$$

ثم نحسب $e_i = Y_i - \hat{Y}_i$

$$e_1 = 4 - 4.13 = -0.13$$

$$\vdots$$

$$e_6 = 11 - 10.35 = 0.47$$

وكما في الجدول (2-6)

جدول (2-6)

جدول حسابات قيم Y المقدرة وأخطاء التقدير

	$\hat{Y}$	e	e^2
	3.56	-.13	.0169
	5.41	.59	.3481
	6.69	.31	.0961
	7.97	-.97	.9409
	9.25	-.25	.0625
	10.53	.47	.2209
Σ	43.41	0.02	1.6854

(٥) اذا $\beta_0 = 0$ فان معادلة الانحدار جبرياً:

$$Y_i = \beta_1 X_i$$

وان مجموع مربعات الخطأ في هذه الحالة هي:

$$S(\hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_i)^2$$

(٦) المعادلات (2-10) و (2-11) تصبح غير ملائمة للتقدير اذا $\beta_0 = 0$ لان دالة الهدف في هذه الحالة هي :

$$S(\hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_i)^2$$

(٧)حيث ان دالة الهدف هي:

$$S(\hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_i)^2$$

فالشرط الضروري:

$$\frac{\partial S}{\partial \hat{\beta}_i} = 2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_i) \cdot (-X_i)$$

ومنها تكون المعادلة الطبيعية:

$$\frac{\partial S}{\partial \hat{\beta}_i} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\therefore \hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{176}{91}$$

$$\therefore \hat{\beta}_1 = 1.934$$

وبذلك فان معادلة الانحدار:

$$\hat{Y}_i = 1.934 X_i$$

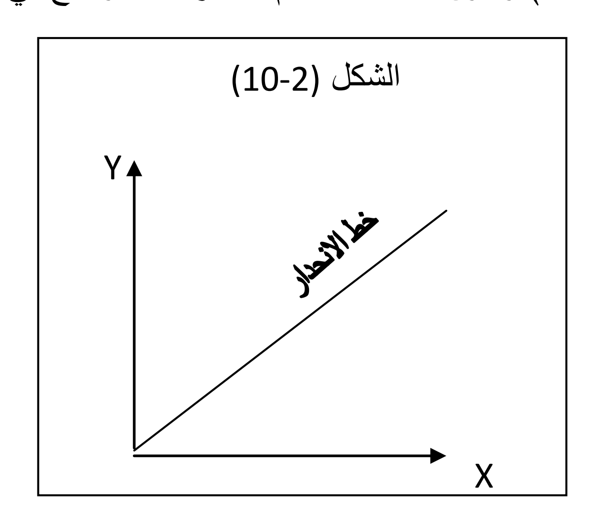
(٨) لرسم الخط المقدر عندما $\beta_0 = 0$

نفرض $X = 1 \quad \Leftarrow \quad \hat{Y} = 1.934$

نفرض $X = 2 \quad \Leftarrow \quad \hat{Y} = 3.868$

ثم نصل النقاط عبر نقطة الأصل (0,0) ويكون الخط المستقيم المقدر كما موضح في الشكل(2-10).

الشكل (2-10)



(٩) وللحصول على بواقي العلاقة نحسب قيم $\hat{Y}$ أولاً من معادلة التقدير.

$$\hat{Y}_i / X_i = 1.934 X_i$$

ثم نحسب الأخطاء

$$e_i = Y_i - \hat{Y}_i$$

وكما موضحة في الجدول (2-7)

جدول (2-7)
القيم الأصلية

XY	X^2	$\hat{Y}$	e	e^2
4	١	1.934	2.061	4.268
12	٤	3.868	2.132	4.545
21	٩	5.802	1.198	1.435
٢٨	١٦	7.736	-.736	0.542
٤٥	٢٥	9.67	-.67	0.449
٦٦	٣٦	11.604	-.604	0.365
∑١٧٦	٩١	40.614		11.604

$$\sum_{i=1}^{6} e_i^2 = 11.604$$

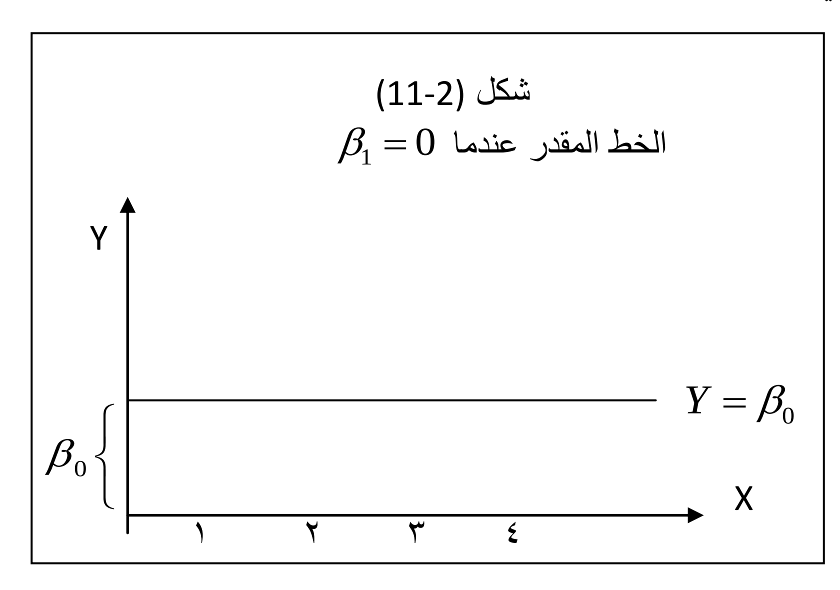
(١٠) $\Sigma e_i^2 = 1.6854$ عندما معادلة التقدير: $\hat{Y}_i = 2.85 + 1.28X_i$

في حين عند معادلة التقدير $\hat{Y}_i = 1.934X_i$ فان $\sum_{i=1}^{6} e_i^2 = 11.604$

اذن الخط المستقيم المقدر يكون أكثر تمثيلاً للبيانات اذا استخدمنا النموذج بوجود المقطع الصادي.

(١١) اذا كانت القيمة الحقيقية لـ $\beta_1 = 0$.

فان الخط المقدر يتبع الشكل التالي:



شكل (2-11)
الخط المقدر عندما $\beta_1 = 0$

كما ان معادلة الانحدار تتبع الصيغة التالية: $E(Y_i / X_i) = \beta_0$

(١٢) من الشكل يتضح إن قيم Y لا تتحدد على وفق قيم X ، فهي لا تتغير مع تغير قيم X بمعنى إن أثر X معدوم في تحديد قيم Y .

(5-2) التنبؤ Prediction :ان نتائج الانحدار يتم تطبيقها في اتجاهين:

(1) اختيار الفرضيات حول سلوك صيغة التقدير والذي بدوره يتكون من خطوتين:
(أ) الاختبارات الإحصائية والتي سيتم مناقشتها في الفصل القادم (الفصل الثالث).
(ب) اختبارات الدرجة الثانية والتي تمثل اختبارات تحقق فروض التحليل والتي يتم تناولها في الباب الثاني.

(٢) والاتجاه الثاني لتطبيق نتائج الانحدار هو التنبؤ بتأثير إحداث معينة على المتغير المستقل.

ويعد التنبؤ من إحدى أهم التطبيقات لنموذج الانحدار. فمثلاً بالرجوع إلى دالة الإنتاج الزراعي المقدرة في المثال (2-1)، يمكن استخدامها للتنبؤ عن حجم الإنتاج الزراعي في حال استخدام سياسة توسعية للمساحات المزروعة. فعند افتراض توسع المساحة المزروعة إلى (X_f) ، فتعوض قيمته في دالة الإنتاج الزراعي المقدرة فنحصل على حجم الإنتاج المستقبلي المناظر لقيمة X_f:

$$Y_f = \beta_0 + \beta_1 X_f + u_f$$

حيث ان u_f قيمة الخطأ العشوائي المستقبلي. والذي لايمكننا التنبؤ به لأنه عشوائي وغير مرتبط باي قيمة سابقة للخطأ العشوائي (بموجب فروض التحليل) $\text{cov}(u_i u_j) = 0$ وكذلك غير مرتبط بقيم المتغير المستقل $\text{cov}(u_i X_j) = 0$.
وعليه يتبين وجود مصدرين لعدم التأكد في القيم التنبؤية:

الاول: عدم معرفتنا بقيم β_1, β_0 ويجب استخدام التقديرات لها $\hat{\beta}_1, \hat{\beta}_0$ ، وهذا يمثل متوسط الاستجابة

$$Y_f^m = \text{E}(Y/X_f) = \beta_0 + \beta_1 X_f$$

اما مصدر عدم التأكد الثاني هو الاثر الذي لايمكن التنبؤ به للخطأ العشوائي .
ويتطلب ذلك الرغبة في الحصول على التنبؤ Y_f الى جانب مقياس لدقة التنبؤ والمتمثل في بناء فترة ثقة له.

(1-5-2) <u>تكوين التنبؤات:</u>

اولاً: تقدير متوسط الاستجابةEstimating the mean prediction: فمع توافر المقدرات $\hat{\beta}_1, \hat{\beta}_0$ والمعتمدة على عينة بحجم (n) للمتغيرين X و Y ، وطالما ان المدة (f) هي مدة مستقبلية فان $(f \rangle n)$ وبظل افتراضات النموذج فان $\hat{\beta}_1$ و $\hat{\beta}_0$ مقدرات غير منحازة . وبذلك يمكن استخدم الصيغة:

$$\hat{Y}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f$$ مقدراً غير متحيز لـ Y_f^m :

$$E(\hat{Y}_f) = E(\hat{\beta}_0 + \hat{\beta}_1 X_f) = E(\hat{\beta}_0) + \left(E(\hat{\beta}_1)\right) X_f$$

$$E(\hat{Y}_f) = \hat{\beta}_0 + \hat{\beta}_1 X_f = Y_f^m \qquad ... \qquad (30-2)$$

وبالرجوع إلى مثال الإنتاج بدلالة المساحة المزروعة وبافتراض توسيع المساحة المزروعة (250)هكتار فان تقدير القيمة المتوقعة للإنتاج:

$$\hat{Y}_f^m = 35.35 + 2.564(250) = 676.35$$ ألف كغم

نلاحظ ان $\hat{Y}_f$ هو تقدير النقطة للقيمة المتوسطة لـ Y_f وهو Y_f^m والمناظرة لـ X_f .

كما ويطلق على هذا النوع من التنبؤ بالتنبؤ ضمن حدود العينة.

ثانياً: تقدير القيمة التنبوئية الجديدة :Estimating the point prediction

القضية ذات الأهمية الرئيسة هي التنبؤ بمشاهدات جديدة أي التنبؤ بـ Y_f ذاتها. وبافتراض قيمة X_f معطاة بمقدورنا التنبؤ بالمستوى المستقبلي لـ Y وهو Y_f. (يعد مفردة جديدة)

وحيث ان: $\hat{Y}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f$ وذلك لان $E(u) = 0$

وببساطة يتم التنبؤ بمستوى Y_f عن طريق التنبؤ بقيمته المتوسطة.

وينشأ عن ذلك خطأ التنبؤ:

$$e_f = Y_f - \hat{Y}_f \qquad ... \qquad (31-2)$$

ويكون متوسط ذلك الخطأ:

$$E(e_f) = E(Y_f) - E(\hat{Y}_f)$$

$$= \beta_0 + \beta_1 X_f - \beta_0 - \beta_1 X_f = 0$$

أسئلة الفصل الثاني:

س1: صحح الخطأ في كل من العبارات التالية:

١- ان الانحراف للقيم المنفردة (Y_i) حول قيمها المتوقعة تمثل خطأ التنبؤ.

٢- ان علاقة الانحدار تتكون من جزأين، الجزء الأول المتوسط الشرطي لـ Y عند مستوى محدد لـ X ، والجزء الثاني هو الجزء المحدد غير الشرطي.

٣- في الانحدار البسيط ميل الخط المستقيم هو المعلمة (β_1) تكون موجبة.

٤- رسم الانتشار هو الشكل الذي تحدد به قيم المتغير المعتمد المقدرة.

٥- معيار طريقة المربعات الصغرى هو مجموع المسافة العمودية لكل نقطة من البيانات عن الخط المقدر تكون أصغر ما يمكن.

٦- الشرط الكافي لتحقيق مبدأ المربعات الصغرى يولد المعادلات الطبيعية.

٧- ان تحليل الانحدار يسعى لتحديد مدى قرب Y المقدرة ب١موجب طريقة التقدير المختارة مع حد الخطأ.

٨- تسمى علاقة الانحدار خطية إذا كانت خطية بدلالة المعلم١ات وكذلك خطية بدلالة المتغيرات.

٩- المتغير العشوائي كروياً يعني متجانس التباين.

١٠- ان مجموع التغيرات الإجمالية في الانحدار يقسم إلى جزأين هما مجموع مربعات الخطأ مضافاً اليها مجموع مربعات متوسط استجابة Y.

١١- عند ثبوت قيمة المتغير Y مقابل تغير واضح في قيم X، فان ذلك دليلاً على معنوية X.

س٢: ناقش فرضيات تحليل الانحدار؟

س3: وضح بالتفصيل ان يكون نموذج الانحدار خطياً ؟

س4: وضح المقصود بالعبارات التالية:

١. المتغير X يفترض ان يكون غير عشوائي؟
٢. علاقة الانحدار خطية بدلالة المعلمات؟
٣. فرضية عدم وجود ارتباط ذاتي بين مشاهدات المتغير العشوائي؟
٤. عدم وجود ارتباط بين ui و Xi ؟
٥. BLUE ؟
٦. المتغير العشوائي الكروي (Spherical disturbance) ؟

س5: أي من العلاقات التالية تحقق فرضية خطية علاقة الانحدار:

$$Y_i = \beta_0 + \beta_1 \sqrt{X_i} + u_i$$

$$Y_i = \frac{\beta_0}{X_i} + \beta_2 X_i^2 + u_i$$

$$\frac{1}{Y_i} = \beta_0 + \beta_1 \ln X_i + u_i$$

$$Y_i = \sqrt{\beta_0} + \beta_1 X_i + u_i$$

$$\ln Y_i = \beta_0 + \beta_1 \ln X_i + u_i$$

$$Y_i = \beta_0 X_i^{\beta_1} e^{u_i}$$

$$Y_i = \beta_0 + \beta_i^2 X_i + u_i$$

س6: وضح بالرسم:

١. تباين الخطأ في الانحدار البسيط يكون متجانساً.

٢. الأنماط المختلفة للمتغير العشوائي (ارتباط ذاتي طردي وعكسي، واستقلال ذاتي).

٣. نص نظرية جاوس – ماركوف.

٤. خط انحدار بسيط ذو ميل سالب ومقطع معدوم.

٥. خط انحدار بسيط ذو ميل معدوم ومقطع موجب.

٦. خط انحدار بسيط ذو ميل موجب ومقطع موجب.

س7:قارن بين طريقتي OLS و ML من حيث (المبدأ ، نتائج التقدير، صفات المقدرات) ؟

س8: اشتق صيغة مناسبة لتباين كل من : معلمات الانحدار β_0 ، β_1 والخطأ للمجتمع بموجب المربعات الصغرى (OLS) وبموجب ML ؟

س9: وضح توزيع المعلمات المقدرة $\hat{\beta}_0$ و $\hat{\beta}_1$؟

س10: اذكر أهم صفات مقدرات المربعات الصغرى ثم بين توزيعها ؟

س11: Y يرتبط بالمتغير X على وفق الصيغة: $Y_i = \beta_0 + \beta_1 e^{2X_i} + u_i$ للبيانات:

X	٧	8	12	19
Y	٢	1	5	7

١. صغ العلاقة بصيغة النموذج الخطي العام؟

٢. اذكر المعادلات الطبيعية للنموذج ؟

س12: الجدول التالي يمثل عدد ساعات المطالعة اليومي (X) والمعدل العام (Y) لعينة من طلبة قسم الإحصاء:

X	4	5	3	6
Y	60	70	75	80

١. أفضل معادلة خطية تربط بين X و Y وتفسير معلماتها ؟
٢. قدر معدل طالب يدرس (2) ساعة ؟
٣. رسم الخط المقدر مع رسم الانتشار ؟

الفصل الثالث

الاستدلال في نموذج المربعات الصغرى Inferences in OLS"" Model

بعد ان تم تقدير معلمات نموذج الانحدار الأساسي ذي المتغيرين، وتحديد صفات المقدرات وتوزيعها في الفصل السابق ننتقل الى مرحلة الاستدلال حول صحة النتائج التقديرية. ويتطلب ذلك مرحلتين:

<u>**المرحلة الأولى:**</u> تتطلب استخدام الآتي:

١- الاستدلال حول المعلمات المقدرة أو القيم التنبؤية ويتمثل ذلك باختبار الفرضيات وكذلك فترات الثقة.

٢- جدول تحليل التباين .Analysis of variance

٣- اختبار حسن التوفيق .Goodness of fit

<u>**المرحلة الثانية:**</u>وتتضمن هذه المرحلة استخدام اختبارات مناسبة للتحقق من ان فرضيات النموذج قد تحققت.

وستختص فقرات الفصل الثالث لما تتطلبه المرحلة الاولى. اما المرحلة الثانية فسيتم عرضها في الباب الثاني.

<u>(1-3)الاستدلال حول المعلمات المقدرة Inference about estimated parameters .</u>

بعد تقدير المعلمات فالسؤال المطروح هل يمكن الركون الى هذه النتائج؟ وما مدى ثقتنا بها.

على سبيل المثال ما مدى ثقتنا في كون المعلمة β_0موجبة في الحقيقة؟ فاذا كان تقديرنا لـ β_0 هو0.001، فهل يكون مقنعاً بالقول ان β_0 موجبة وليست صفراً.

واذا افترضنا ان $\beta_1 = 0.9$ ، فهل ان تقديرنا لـ β_1 $(\hat{\beta}_1 = 0.72)$ غير متسق مع الافتراض$\beta_1 = 0.9$؟

فقد يكون التفاوت بين التقدير والقيمة الحقيقية نشأ من حجم العينة التي تم استخدامها في التقدير. فاذا كانت درجة مادة الاحصاء في المرحلة الاولى من قسم الاحصاء في جامعة البصرة هي (69)، فلا ينبغي ان نتوقع ان الدرجة المتوسطة لمجموعة عشوائية من (10) أفراد مثلاً يكون (69) بالضبط.

في هذه الفقرة نهتم بدراسة التناسق بين تقديرات المعلمات والفرضيات المسبقة، وكذلك بناء فترة ثقة حول تقديرنا، لتمكننا من استخدام تقدير النقطة لايجاد التقدير بالفترة.

(1-1-3)تفسير اختبار الفرضيات Interpretation of testing hypothese

ان منهج الاختبار يتلخص بوضع فرضية العدم (H_0) وفرضية بديلة (H_1) ونسعى لقبول فرضية العدم أو رفضها على اساس الفرق بين القيمة المفترضة للمعلمة وتقديرنا لها. فاذا كان الفرق بين ($\hat{\beta}$) وبين القيمة المفترضة (β) أي (قيمة المعلمة في المجتمع الذي سحبت منه العينة) كبيراً جداً فيكون القرار برفض فرضية العدم، وفرضية العدم تحدد فيها القيم التي يعتقد الباحث بانها لا تعبر عن القيمة الحقيقية للمعلمة. ولتحديد مقدار الفرق الكبير، لابد من اختيار مستوى دلالة معين.

(2-1-3)مستوى الدلالة Level of significance.

يطلق عليه الخطأ من النوع الأول (Type Ierror) وهو يرمز لاحتمال الوقوع بهذا الخطأ عند رفض فرضية العدم H_0 عندما تكون هذه الفرضية صحيحة في الحقيقة. ويرمز له بالرمز (α) ويسمى أيضاً مستوى المعنوية.

وهناك خطأ آخر محتمل ان نقع به هو ان نقبل H_0 حتى اذا كانت الفرضية البديلة هي الفرضية الصحيحة. ويطلق على هذا النوع من الخطأ (الخطأ من النوع الثاني) (Type Π error) ويرمز له بالرمز(β). وجدير بالذكر ان تخفيض احتمال الوقوع في الخطأ من النوع الأول سيزيد في الوقت نفسه احتمال الوقوع في الخطأ من النوع الثاني. وفي العادة يتم الاختبار باستخدام α عند (0.05) أو (0.01). وضرورة التأكيد في هذا الصدد ان أي من الخطأين قد يؤدي إلى القيام بعمل له نتائج غير مرغوب فيها. فهناك خسارة مرتبطة بكل نوع من الأخطاء ونختار الخسارة الأقل.

(3-1-3)الفرضيات حول المعلمات المقدرة.parametersHypotheses about estimated

في الغالب ان الظاهرة المطلوب دراستها هي التي تقترح طبيعة فرضية العدم والفرضية البديلة لكلا المعلمتين β_0 و β_1. وسيتم تحديد أهم انواع الفرضيات ولكل معلمة من المعلمات على وفق الآتي:

١– اذا اردنا اثبات ان الميل لخط الانحدار β_1 او الحد الثابت β_0 لا يساوي صفراً فتكون الفرضية:

$H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

أو: . . . (1-3)

$H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$

وهكذا يمكن تعميمه لاي قيمة غير الصفر.

ويسمى الاختبار بهذه الحالة اختبار بجانبين (Two tail test)

٢– اذا كنا نسعى الى اثبات ان ميل انحدار المجتمع موجباً او ان الحد الثابت موجباً فتكون الفرضية:

$H_0: \beta_1 \leq 0$ vs. $H_1: \beta_1 > 0$

(2-3) . . . أو:

$H_0: \beta_0 \leq 0$ vs. $H_1: \beta_0 >$

ويسمى الاختبار في هذه الحالة اختبار بجانب واحد (One – tail test)

٣- أما اذا أردنا اثبات ان ميل انحدار المجتمع سالباً او ان الحد الثابت سالباً فان الفرضية تصاغ على وفق الآتي:

$H_0: \beta_1 \geq 0$ vs. $H_1: \beta_1 < 0$

(3-3) . . . أو:

$H_0: \beta_0 \geq 0$ vs. $H_1: \beta_0 < 0$

وهكذا يمكن تعميم الحالات الثلاث السابقة لاي قيمة معلومة (غير الصفر) وكالآتي:

الحالة الاولى:

$H_0: \beta_1 = (\beta_1)_0$ vs. $H_1: \beta_1 \neq (\beta_1)_0$

(1-3)' . . . أو:

$H_0: \beta_0 = (\beta_0)_0$ vs. $H_1: \beta_0 \neq (\beta_0)_0$

الحالة الثانية:

$H_0: \beta_1 \leq (\beta_1)_0$ vs. $H_1: \beta_1 > (\beta_1)_0$

(2-3)' . . . أو:

$H_0: \beta_0 \leq (\beta_0)_0$ vs. $H_1: \beta_0 > (\beta_0)_0$

الحالة الثالثة:

$H_0: \beta_1 \geq (\beta_1)_0$ vs. $H_1: \beta_1 < (\beta_1)_0$

(3-3)' . . . أو:

$H_0: \beta_0 \geq (\beta_0)_0$ vs. $H_1: \beta_0 < (\beta_0)_0$

والجدير بالذكر ان صياغة الفرضيات واختيار حجم الخطأ من النوع الاول(α) ينبغي ان تسبق تحليل البيانات والحصول على المقدرات.

(4-1-3)الاحصاءة المستخدمة للاختبار testStatistic forUsed .

لإجراء اختبار الفرضيات المؤشرة في الحالات الثلاث المؤشرة في الفقرة السابقة ينبغي علينا اولاً أن نستذكر خصائص المتغير العشوائي u_t التي تم افتراضها، فهي متغيرات عشوائية ذات قيم متوقعة صفرية ، ولها قيمة ثابتة للتباين ولها تغايرات صفرية كما انها مستقلة عن المتغيرات الموجودة في الطرف الأيمن من المعادلة (وهنا هو المتغير المستقل X). كما ان الأخطاء العشوائية موزعة توزيعاً طبيعياً أي

ان دالتها الاحتمالية هو المنحنى الطبيعي* وبذلك يمكن تحديده بواسطة الوسط الحسابي والتباين. والتي يمكن إدراجها كالآتي:

$$u_t \sim N(0, \sigma^2)$$

وحيث ان: $Y = \beta_0 + \beta_1 X + u$

فان: $Y \sim N(\beta_0 + \beta_1 X, \sigma^2)$

وكما أوضحنا في الفصل الثاني ان $\hat{\beta}_0$ و $\hat{\beta}_1$ توليفات خطية بدلالة الأخطاء العشوائية u_1 ، . . . . ، u_n لقيم معطاة للمتغير المستقل X . وهي مقدرات غير متحيزة، وتم اشتقاق صيغاً لتبايناتها وبذلك يمكن ان نكتب توزيع هذه المقدرات كالآتي:

$$\hat{\beta}_0 \sim N\left(\beta_0 ,\ \sigma^2 \frac{\Sigma X^2}{n\Sigma x^2}\right)$$

وكذلك

$$\hat{\beta}_1 \sim N\left(\beta_1 ,\ \frac{\sigma^2}{\Sigma x^2}\right)$$

وبذلك يمكننا استخدام طرق الإحصاء لاختبار الفرضيات المرتبطة بـ β_0 و β_1 ومبادئ الإحصاء الأولية التي يمكننا من استخدام توزيع Z الطبيعي المعياري وجداوله الخاصة. واستخدام الاحصاءة Z يتطلب ان تكون قيمة تباين المجتمع معلوماً. وحيث ان تباين المعلمات المقدرة $\hat{\beta}_0$ و $\hat{\beta}_1$ تعتمد على تباين الخطأ σ^2 وهي قيمة غير معلومة لذلك فان تباين المعلمات المقدرة $\hat{\beta}_0$ و $\hat{\beta}_1$ هي الأخرى غير معلومة. ولقد تم استخدام المقدر $\hat{\sigma}^2$ من أجل الحصول على تباينات المقدرات حيث

$$\hat{\sigma}^2 = \frac{\Sigma e_i^2}{(n-2)}$$

ومع هذا التعديل فليس بالإمكان استخدام المنحنى الطبيعي لاختبار الفرضيات وإنما يتم اعتماد توزيع ستيودنت t** بدرجات حرية (n-2) . ويتم استخدام الجداول الخاصة بتوزيع t ولدرجات حرية (n-2) لاستخراج القيم النظرية. وبذلك فان الاحصاءة المستخدمة هي:

* ان كثير من النتائج المذكورة في هذا الكتاب لا تتطلب من الناحية الفنية لهذا الافتراض.

** يشبه توزيع t التوزيع الطبيعي. ومع كبر حجم العينة (n) فان التوزيع سوف يؤول الى التوزيع الطبيعي.

$$t_{\hat{\beta}_0} = \frac{\hat{\beta}_0 - (\beta_0)_0}{s.e(\hat{\beta}_0)}$$
$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - (\beta_1)_0}{s.e(\hat{\beta}_1)} \quad \ldots \quad (4-3)$$

لاختبار الحالات الثلاث المذكورة في المعادلات $(1-3)'$ و $(2-3)'$ و $(3-3)'$.

وتسمى القيم في المعادلة (3-4) قيم t العملية (المحسوبة) وتقارن هذه القيم العملية مع النظرية والتي يتم استخراجها من جداول خاصة بتوزيع t ولدرجات حرية (n-2) وبموجب مستوى معنوية مختار مسبقاً ويرمز له: $t_{c(n-2,\gamma)}$

مع ملاحظة: γ : هو الاحتمال الذي يكون متغير t بدرجات حرية (n-2) أقل من t_c

$\gamma = \frac{\alpha}{2}$: في حالة اختبار الطرفين.

$\gamma = \alpha$: في حالة اختبار الطرف الواحد.

(5-1-3)نسبة t : قاعدة للحساب .t– Ratio

ان الاستنتاج الذي يمكن الوصول اليه يعتمد على حالة الاختبار، ويمكن تحديد الحالات التالية:

الحالة الأولى: في حالة الاختبار ذي الطرفين، اذا كانت قيمة المعلمة المقدرة أكبر من ضعف حجم الخطأ المعياري المقدر، فان الاستنتاج يكون اختلاف المعلمة المقدرة معنوياً عن الصفر وعند مستوى معنوية 5% بمعنى رفض فرضية العدم المذكورة في المعادلة (3-1) .

أما اذا كانت قيمة المعلمة المقدرة اكبر من ثلاثة أمثال حجم الخطأ المعياري المقدر فان هذه المعلمة مختلفة عن الصفر عند مستوى معنوية 1% .

وبشكل عام المخطط التالي يوضح منطقة القبول والرفض للفرضية بجانبين وبافتراض مستوى معنوية α %:



المساحة المضللة تمثل منطقة الرفض Rejection Region ، فعندما تقع القيمة المحسوبة لـ t في منطقة الرفض يكون القرار برفض H_0 أي عدم رفض H_1 . والعكس صحيح.

بمعنى :
$$\left| t_{\hat{\beta}} \right| > t_{c(n-2,\frac{\alpha}{2})}$$

يكون القرار رفض H_0.

مثال (3-1): لوحة البيانات عن دراسة لفحص العلاقة بين الخبرة والاجر لمجموعة افراد في حرفة معينة ومنطقة جغرافية معينة، حيث ان الخبرة متمثلة بعدد سنوات الخدمة (X) والاجر (Y) بالف دولار ولعينة مكونة من (16) حرفي.

$$\Sigma Y = 750 \quad , \quad \Sigma X = 245 \quad , \quad \Sigma XY = 13210 \quad , \quad \Sigma X^2 = 5291 \quad , \quad \Sigma Y^2 = 38080$$

م/ 1- قدر علاقة انحدار Y بدلالة X للعينة.

2- فسر المعلمات المقدرة

3- اختبر معنوية المقدرات

الحل: (1) تقدير معادلة الانحدار :

$$\hat{\beta}_1 = \frac{\Sigma XY - \frac{1}{n}(\Sigma X)(\Sigma Y)}{\Sigma X^2 - \frac{1}{n}(\Sigma X)^2} = \frac{S_{XY}}{S_{XX}}$$

$$S_{XY} = 13210 - \frac{1}{16}(245)(750)$$

$$= 13210 - 11484.38$$
$$= 1725.62$$

$$S_{XX} = 5291 - \frac{1}{16}(245)^2$$

$$= 5291 - 3755.238$$
$$= 1539.44$$

$$\hat{\beta}_1 = \frac{1725.62}{1539.44} = 1.121$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$
$$= 46.88 - (1.121)(15.3125)$$
$$= 29.71$$

اذن معادلة الانحدار المقدرة: $\hat{Y} = 29.71 + 1.121X$

(2) تفسير المعلمات:

$\hat{\beta}_0 = 29.71$: تمثل متوسط الأجر عند أول التعيين أي أجر العامل في بداية التعيين.

$\hat{\beta}_1 = 1.121$: اثر الخبرة في أجر العامل. وهي موجبة اذن زيادة سنوات الخبرة سنة واحدة تسهم في زيادة الأجر لدى الأفراد بمقدار (1.121) ألف دولار.

(3) اختبار معنوية المقدرات: بالنسبة للمعلمة $\hat{\beta}_1$:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

نسبة t:

$$t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)}$$

$$s.e(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = \sqrt{\frac{\hat{\sigma}^2}{\Sigma x^2}} = \sqrt{\frac{\hat{\sigma}^2}{S_{XX}}}$$

حيث ان: $\hat{\sigma}^2 = \dfrac{\Sigma e^2}{n-2}$

$$\Sigma e^2 = \Sigma y^2 - \Sigma \hat{y}^2$$

و $\Sigma \hat{y}^2 = ESS = \hat{\beta}_1 S_{xy} = (1.121)(1725.62) = 1934.42$

$$S_{yy} = TSS = \Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 38080 - \frac{(750)^2}{16} = 2923.75$$

$$\therefore \Sigma e^2 = RSS = TSS - ESS$$
$$= 2923.75 - 1934.42$$
$$= 989.33$$

$$\hat{\sigma}^2 = \frac{989.33}{14} = 70.67$$

$$\therefore \text{var}(\hat{\beta}_1) = \frac{70.67}{1535.76} = 0.046$$

$$\therefore s.e(\hat{\beta}_1) == \sqrt{0.046} = 0.214$$

نسبة t للمعلمة $\hat{\beta}_1$ هي:

$$t^*_{\hat{\beta}_1} = \frac{1.121}{0.214}$$

$$= 5.238$$

نختار مستوى دلالة 5% ونستخرج القيمة النظرية للنسبة t من جداول t بدرجات حرية (n-2) وتحت مستوى دلالة 0.025 لان الاختبار ذو طرفين:

$$t_{c(14,0.025)} = 2.145$$

القرار: لان $t_{c(14,0.025)} \; \langle \; t^*_{\hat{\beta}_1} \Leftarrow$ نرفض H_0 أي ان المعلمة β_1 في المجتمع تختلف عن الصفر. أو ان المتغير X مهم في تحديد التغيرات في المتغير Y. بعبارة أخرى ان سنوات الخبرة مهمة في تحديد قيمة الأجر السنوي.

وبالطريقة نفسها يتم اختبار الفرضية:

$H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$

$$t^*_{\hat{\beta}_0} = \frac{\hat{\beta}_0}{s.e(\hat{\beta}_0)} = \frac{\hat{\beta}_0}{\sqrt{\text{var}(\hat{\beta}_0)}}$$

$$\text{var}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \Sigma X^2}{n\ S_{XX}} = 70.67 \cdot \frac{5291}{16(1539.44)} = \frac{373914.97}{24631.04}$$

$$= 15.18$$

$$s.e(\hat{\beta}_0) = 3.896$$

$$t^*_{\hat{\beta}_0} = \frac{29.71}{3.896} = 7.626$$

$$t_{c(14,0.025)} = 2.145$$

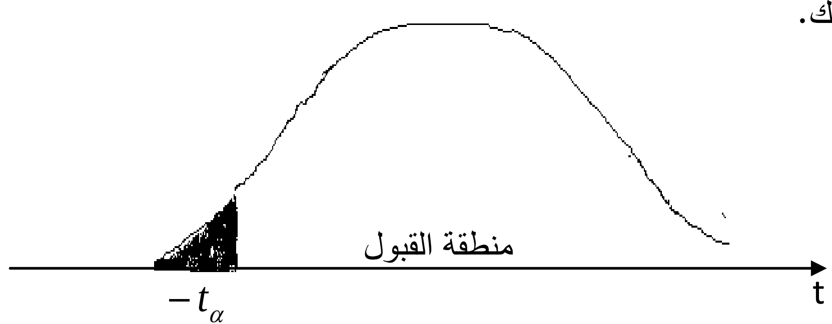
اذن نرفض H_0 أي ان المعلمة β_0 معنوية إحصائياً.

<u>الحالة الثانية</u>: في حالة الاختبار بذيل واحد (جانب واحد)، نكون نسبة t من المعادلة (4-3) ونقارنها بـ $t_{c(n-2,\alpha)}$.

اذا كانت القيمة الحسابية ضمن منطقة القبول $t^*_{\hat{\beta}_i} \geq -t_{c(n-2,\alpha)}$.

فان القرار يكون بقبول الفرضية H_0 والعكس صحيح.

والمخطط يوضح ذلك.



المنطقة المضللة تمثل منطقة الرفض.

مثال (3-2) استخدم معلومات المثال السابق نفسها.

اختبر: $H_0: \beta_1 > 1$

ضد

$H_1: \beta_1 \le 1$

قيمة t المحسوبة للمعلمة $\hat{\beta}_1$ هي:

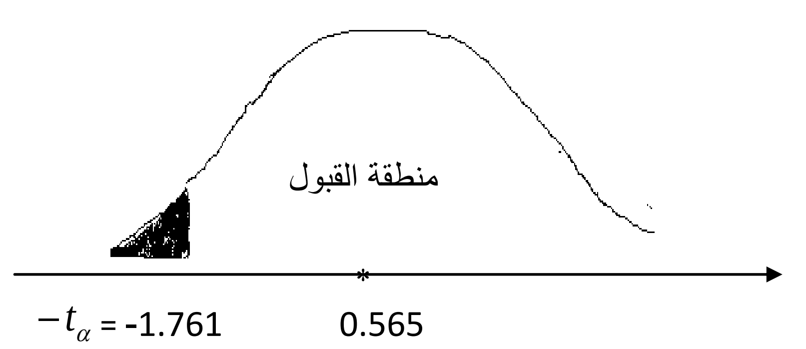
$$t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - 1}{s.e(\hat{\beta}_1)}$$

$$= \frac{0.121}{0.214} = 0.565$$

الاختبار من طرف واحد باستخدام مستوى دلالة %5 مثلاً:

$t_{c\,(14,0.05)} = 1.761$

وبإتباع المخطط



فان القيمة المحسوبة لـ t تقع ضمن مجال القبول

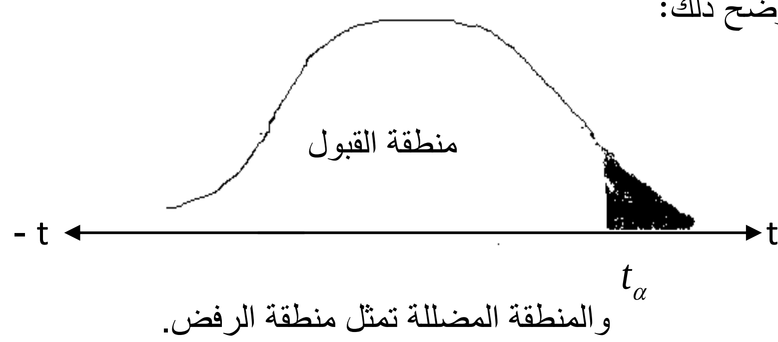
فالقرار يكون بقبول فرضية العدم بان β_1 اكبر من واحد.

الحالة الثالثة: بعد تكوين نسبة t من العلاقة (3-4) تتم مقارنتها بـ t النظرية بدرجات حرية (n-2) ومستوى دلالة % α.

اذا كانت القيمة الحسابية ضمن منطقة القبول: $t^*_{\hat{\beta}} \leq t_{c(n-2,\alpha)}$

فيكون القرار بقبول فرضية العدم والعكس صحيح.

والمخطط يوضح ذلك:



والمنطقة المضللة تمثل منطقة الرفض.

مثال (3-3) : عينة بعدد 15 مشاهدة، اذا اعطيت معادلة التقدير:

$$\hat{Y}_i = 0.71 + 0.2X_i$$
$$s.e: \quad (0.27)$$

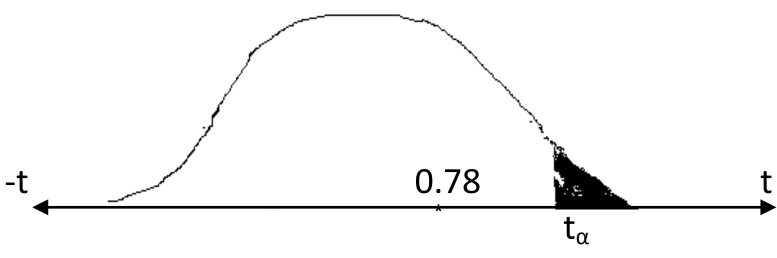
اختبر الفرضية: $H_0: \beta_1 < 0$

$H_1: \beta_1 \geq 0$

$$t^*_{\hat{\beta}_1} = \frac{0.211}{0.27} = 0.78$$

باستخدام مستوى دلالة % 5

$t_{c\,(13,0.05)} = 1.771$



1.1771

اذن القيمة المحسوبة تقع ضمن مجال القبول

القرار: ان معلمة الانحدار سالبة.

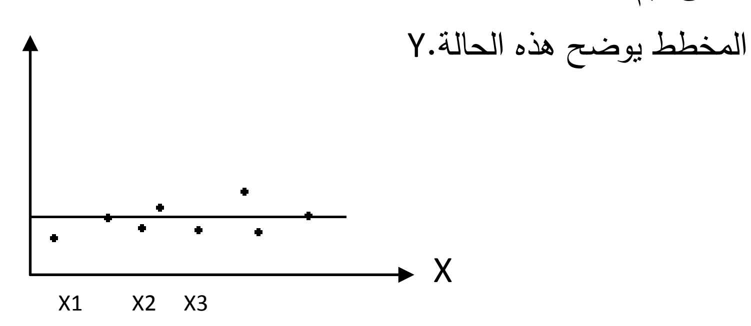
يتضح من الامثلة المذكورة ان اختبارات المعنوية التي تخص المعلمة β_1 هي اكثر أهمية من الاختبارات التي تخص معلمة الحد الثابت (β_0) . اذ ان المعلمة β_1 والتي تمثل معلمة الانحدار فان اختبار معنويتها

تدل على اختبار معلمة الانحدار واختبار للعلاقة بين متغير الاستجابة Y مع المتغير المستقل X فان قبول الفرضية $H_0: \beta_1 = 0$

يشير إلى أحد الأمرين:

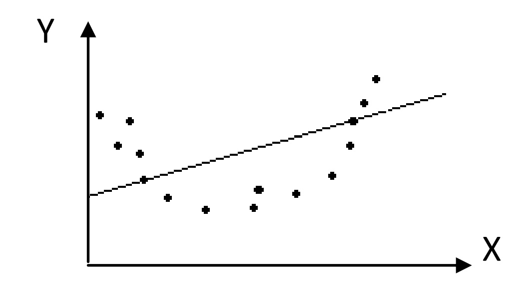
1- ان المتغير المستقل X ليس له تأثير معنوي حقيقي على الاستجابة وبذلك فان $Y = \bar{Y}$ لأي قيمة من قيم X .

المخطط يوضح هذه الحالة.Y



2- أو ان نموذج الانحدار $Y = \beta_0 + \beta_1 X + u$ غير ملائم لتمثيل البيانات أي ان العلاقة الحقيقية بينX و Y هي غير خطية.

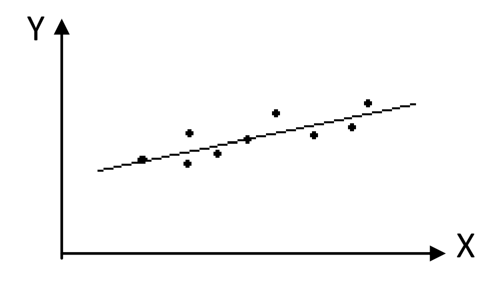
المخطط يوضح ذلك.



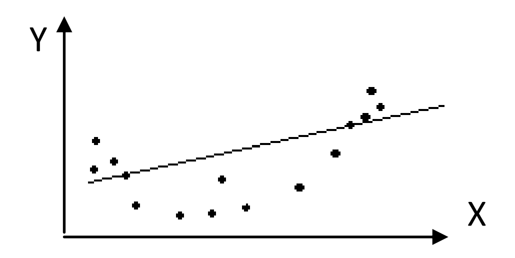
أما رفض الفرضية: $H_0: \beta_1 = 0$

فانه يشير إلى ان X له أهمية في تفسير التغيرات في المتغير Y وذلك يؤكد أحد أمرين:

١- ان العلاقة بين X و Y قد مثلت البيانات تمثيلاً مناسباً. وبذلك فان العلاقة الخطية مناسبة لتمثيل المجتمع. والمخطط يوضح ذلك.



٢- أو بالرغم من وجود تأثير خطي لـ X على Y فان نموذجاً بدرجات أعلى يكون أفضل لتمثيل البيانات، والمخطط يوضح ذلك.



(2-3)حدود الثقة للمعلمات المقدرة.parametersinterval of the estimatedConfidence

ان تقدير المعلمات β_0 و β_1 والتي تمت في الفصل الثاني فهي تمثل تقدير نقطة أو قيمة واحدة (Point estimation) وهي $\hat{\beta}_0$ و $\hat{\beta}_1$ على التوالي واعتماداً على بيانات العينة.

والباحث يرغب في ان يحدد فترة حول قيمة المعلمة الحقيقية (β_0 أو β_1) لتشعره بدرجة من الثقة. وهو ما يسمى بحدود الثقة. فهو عملية إيجاد القيمة الحقيقية للمعلمة بين حدي ثقة أدنى واعلى وباحتمال مقداره ($1-\alpha$)وذلك باستخدام تقديرات النقطة للمعلمات.

الحد الأدنى (L)Lower limit

الحد الأعلى (U)Upper limit

ودون الخوض بتفاصيل أخرى فان تقدير فترة الثقة للمعلمات المقدرة يعتمد على التباين للمعلمات المقدرة وعدد المشاهدات للعينة التي يتم اختيارها لتمثيل المجتمع الى جانب مستوى المعنوية α.

وحيث ان:

$$\frac{\hat{\beta}_i - \beta_i}{s.e(\hat{\beta}_i)} \sim t$$

لذا فان فترة الثقة باحتمال 100($1-\alpha$) للمعلمة β_i في معادلة الانحدار هي الفترة المغلقة:

$$\left[-t_{c(n-2,\frac{\alpha}{2})} \leq \frac{\hat{\beta}_i - \beta_i}{s.e(\hat{\beta}_i)} \leq t_{c(n-2,\frac{\alpha}{2})} \right]$$

أو بعد التبسيط:

$$\left[\hat{\beta}_i - s.e(\hat{\beta}_i) \cdot t_{c(n-2,\frac{\alpha}{2})} \leq \beta_i \leq \hat{\beta}_i + s.e(\hat{\beta}_i) \cdot t_{c(n-2,\frac{\alpha}{2})} \right] \quad ... \quad (5-3)$$

كما ويمكن استخدام حدود الثقة كأسلوب لاختبار معنوية معلمة معينة فاذا كانت القيمة المختبر حولها $(\beta_i)_0$ تقع ضمن حدود الثقة فذلك يدلل على قبول فرضية العدم والعكس صحيح.

مثال (3-4) : استخدم معلومات المثال (3-1) نفسها ص50.

حدود الثقة لمعلمة الانحدار β_1 باحتمال 95% هي:

الحد الأدنى:

$$L = \hat{\beta}_1 - s.e(\hat{\beta}_1) \cdot t_{c(14,0.025)}$$
$$= 1.121 - 0.214(2.145) = 0.662$$

الحد الأعلى:

$$U = \hat{\beta}_1 + s.e(\hat{\beta}_1) \cdot t_{c(14,0.025)}$$
$$= 1.121 + 0.214(2.145) = 1.58$$

اذن المعلمة β_1 الحقيقية نجدها ضمن الفترة (L,U) باحتمال 95%.

وحيث ان قيمة الصفر غير مضمنة في هذه الفترة.

0 0.662 1.58

اذن يمكن استخدام فترة الثقة أيضاً لاختبار معنوية المعلمة، وبهذا فان β_1 تعد معنوية باستخدام مستوى معنوية 5% وذلك لأن الصفر غير مضمن في فترة الثقة للمعلمة باحتمال 95%.

أما حدود الثقة باحتمال 95% لمعلمة المقطع الصادي فهي:

الحد الأدنى:

$$L = \hat{\beta}_0 - s.e(\hat{\beta}_0) \cdot t_{c(14,0.025)}$$
$$= 29.71 - 3.896(2.145) = 21.35$$

الحد الأعلى:

$$U = \hat{\beta}_0 + s.e(\hat{\beta}_0) \cdot t_{c(14,0.025)}$$
$$= 29.71 + 3.896(2.145) = 38.066$$

أي ان قيمة المقطع الصادي β_0 للمجتمع نجدها في المجال (38.066 , 21.35) وباحتمال 95% .

ويمكن ملاحظة ان قيمة الصفر غير مضمنة في تلك الفترة لذا نستنتج بان المعلمة β_0تختلف معنوياً عن الصفر.

ان الربط الجوهري بين (مجال الثقة) و (اختبار المعنوية) يمكن تلخيصه على وفق الاتي:

– في حالة مجال الثقة نحاول تحديد المجال او المدى الذي يمتلك احتمالية محدودة لاحتواء القيمة الحقيقية للمعلمة والتي تكون غير معلومة.

– اما في حالة اختبار المعنوية فاننا نفترض قيمة معينة للمعلمة ثم نحاول ان نرى فيما اذا كانت القيمة المقدرة $(\hat{\beta}_i)$ محصورة في مجال معقول حول القيمة الافتراضية

(3-3)الاستدلال حول متوسط الاستجابة الحقيقي عند مستوى معلوم للمتغير X. Inference about mean prediction at given value of X

في الفصل الثاني تم التطرق إلى أحدى أهم التطبيقات لنموذج الانحدار وهو تكوين التنبؤات والتي تم تصنيفها إلى تنبؤات ضمن حدود العينة أو تقدير متوسط الاستجابة والآخر هو تقديرات للقيمة التنبؤية الجديدة. وفي هذه الفقرة سيتم تحديد توزيع المعاينة لكل من متوسط الاستجابة وكذلك للقيمة التنبؤية الجديدة من اجل الحصول على فترة ثقة أو اختبار الفرضيات حولهما.

(1-3-3)التوزيع الاحتمالي لـ $\hat{Y}_f$ distribution for mean predictionprobability.

نفترض ان X_f هي قيمة من القيم الحقيقية للمتغير X . ونرغب في عمل تقدير فترة ثقة لمتوسط الاستجابة $\mathrm{E}(Y/X = X_f)$ ، ومع توافر المقدرات $\hat{\beta}_0$ و $\hat{\beta}_1$ بالاعتماد على عينة بحجم (n) من مشاهدات (X) و (Y)، وحيث ان $\hat{\beta}_0$ و $\hat{\beta}_1$ هي مقدرات غير متحيزة لـ β_0 و β_1 ، لذا فان الصيغة $\hat{Y}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f$ هي مقدر غير متحيز لـ Y_fبعبارة اخرى:

$$\mathrm{E}(Y/X = X_f) = \hat{Y}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f$$

والقيمة المتوسطة لـ $\hat{Y}$ هي Y_f . أما نوع التوزيع لـ Y_f فهو التوزيع الطبيعي، لاي قيمة من قيم X_fالمعطاة، لانها تركيب خطي بدلالة $\hat{\beta}_0$ و $\hat{\beta}_1$ والتي لها توزيع طبيعي.

ويبقى تحديد تباين $\hat{Y}_f$:

$$\begin{aligned}
\hat{Y}_f &= \hat{\beta}_0 + \hat{\beta}_1 X_f \\
&= \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_f \\
&= \bar{Y} + \hat{\beta}_1 (X_f - \bar{X}) \\
\mathrm{var}(\hat{Y}_f) &= \mathrm{var}\left(\bar{Y} + \hat{\beta}_1 (X_f - \bar{X})\right) \\
&= \mathrm{var}(\bar{Y}) + \mathrm{var}\left(\hat{\beta}_1 (X_f - \bar{X})\right) + 2\,\mathrm{cov}\left(\bar{Y}, \hat{\beta}_1 (X_f - \bar{X})\right)
\end{aligned}$$

وحيث ان $\hat{\beta}_1$ لا تعتمد على $\bar{Y} \Leftarrow$ $\quad \mathrm{cov}\left(\bar{Y}, \hat{\beta}_1 (X_f - \bar{X})\right) = 0$

$$\Rightarrow \mathrm{var}(\hat{Y}_f) = \frac{\sigma^2}{n} + (X_f - \bar{X})^2\,\mathrm{var}(\hat{\beta}_1)$$

$$= \frac{\sigma^2}{n} + (X_f - \bar{X})^2 \cdot \frac{\sigma^2}{S_{XX}} \qquad \ldots \qquad (6\text{-}3)$$

$$\sigma^2_{\hat{Y}_f} = \sigma^2 \left[\frac{1}{n} + \frac{(X_f - \bar{X})^2}{S_{XX}} \right]$$

ويمكن تلخيص ذلك:

$$\hat{Y}_f \sim N\left[\beta_0 + \beta_1 X_f \;\; ; \;\; \sigma^2 \left[\frac{1}{n} + \frac{(X_f - \bar{X})^2}{\Sigma x_i^2} \right] \right] \qquad \ldots \qquad (7\text{-}3)$$

وحيث ان σ^2 غير معلومة ويمكن تقديرها على وفق:

$$\hat{\sigma}^2 = \frac{\Sigma e_i^2}{n-2}$$

وعليه فان نسبة t لمتوسط الاستجابة:

$$t = \frac{\hat{Y}_f - \mathrm{E}(Y/X = X_f)}{\sqrt{\mathrm{var} \;\; \hat{Y}_f}} \sim t_{(n-2)} \qquad \ldots \qquad (8\text{-}3)$$

(2-3-3) حدود الثقة لمتوسط الاستجابة. Confidence interval for mean prediction

وبذلك فان فترة الثقة باحتمال (1 – α) لـ $\hat{Y}_f$ يحسب على وفق الآتي:

$$\hat{Y}_f - t_{c(n-2,\frac{\alpha}{2})} \cdot \hat{\sigma}_{\hat{Y}_f} \le \mathrm{E}(Y/X = X_f) \le \hat{Y}_f + t_{c(n-2,\frac{\alpha}{2})} \cdot \hat{\sigma}_{\hat{Y}_f} \qquad \ldots \qquad (9\text{-}3)$$

مثال (5-3): بالعودة الى المثال (1-3) لدراسة العلاقة بين الخبرة والأجر. افترض ان عدد سنوات الخدمة لأحد أفراد المجتمع $X_f = 10$ فان أجره يحدد على وفق:

$$\hat{Y}_{10} = 29.71 + 1.121(10)$$
$$= 40.92$$

(40.92) ألف دولار أجرة الفرد الذي خدمته (10) سنوات.

$$\Sigma e^2 = 989.33 \;\; ; \;\; \hat{\sigma}_u^2 = 70.67 \;\; ; \;\; \bar{X} = 15.3125 \;\; ; \;\; \Sigma x^2 = 1539.44$$

$$s.e(\hat{Y}_{10}) = \sqrt{\frac{\Sigma e^2}{n-2}\left(\frac{1}{n} + \frac{(X_f - \bar{X})^2}{S_{XX}} \right)}$$

$$\mathrm{var}(\hat{Y}_{10}) = 70.67\left(\frac{1}{16} + \frac{(10 - 15.3125)^2}{1539.44} \right)$$

$$= 70.67\left(\frac{1}{16} + 0.01833\right) = 70.67 \quad \cdot \quad (0.08083)$$

$$= 5.7122$$

$$s.e(\hat{Y}_{10}) = 2.3900$$

∴مجال الثقة لمتوسط الاستجابة عندما $X_f = 10$ وباحتمال 95% هو:

40.92 ±(2.39) ·(2.145) ≡ 40.92 ± 5.12655

أي E (Y(35.79 345 ≤/X_f = 10) ≤ 46.046)

ويمكن بالطريقة نفسها إيجاد فترة ثقة (E (Y/X=X_f ولكل قيمة من قيم X الأصلية.

مثال (6-3): حصل باحث على الخلاصة الإحصائية التالية:

$$\hat{Y}_i = 2.9 + 1.35X_i \quad , \quad i = 1, \ . \ . \ . \ ,15$$

$$S_{YY} = 290.5 \quad , \quad S_{XX} = 112.4 \quad , \quad \overline{X} = 20$$

اختبر معنوية متوسط الاستجابة عند $X_f = 10$ باستخدام مستوى معنوية 5%.

الحل:

$$\hat{Y}_i = 2.9 + 1.35(10) = 16.4$$

$$ESS = \hat{\beta}_1^2 S_{XX} = (1.35)^2(112.4) = 204.849$$

$$RSS = S_{YY} - ESS = 85.651$$

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{85.651}{13} = 6.588$$

$$\text{var}(\hat{Y}_{10}) = 6.588\left[\frac{1}{15} + \frac{(10-20)^2}{112.4}\right]$$

$$= 6.588(0.067 + 0.89) = 6.305$$

$$s.e(\hat{Y}_{10}) = 2.511$$

$$\text{H}_0 : \text{E}(Y/X_f = 10) = 0 \quad v.s \quad \text{H}_1 : \text{E}(Y/X_f = 10) \neq 0$$

$$t^*_{\hat{y}_{10}} = \frac{16.4}{2.511} = 6.531$$

تقارن نسبة t المحسوبة مع قيمة t الجدولية عند مستوى دلالة 0.025 ودرجات حرية (13)

$$t_{c(13,0.025)} = 2.16$$

وبما ان القيمة المحسوبة (العملية) لنسبة t اكبر من القيمة الجدولية (النظرية) نرفض فرضية العدم.

بمعنى ان متوسط الاستجابة عند $X_f = 10$ للمجتمع تختلف معنوياً عن الصفر.

(4-3)التوزيع الاحتمالي وحدود الثقة لـ Y_f التنبؤية الجديدة: Probability distribution and level of significance for individual prediction

ان التنبؤ المستقبلي لـ $Y(Y_f)$ ، يكون متطابقاً مع المقدر $\hat{Y}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f$

الذي يمثل متوسط الاستجابة عندما $X = X_f$.

غير ان هناك فرقاً بين تقدير متوسط الاستجابة $E(Y/X = X_f)$ وبين التنبؤ بالاستجابة الجديدة $(Y_{f\ new})$. في الأول يتم تقدير متوسط التوزيع إلى y فبتكرار المعاينة ستتولد لدينا عدة معادلات انحدار تقديرية وبتقدير متوسط التوزيع إلى Y يتم تقدير متوسط الاستجابة. أما في حالة التنبؤ بالقيمة الجديدة فإننا نتنبأ بنتيجة مفردة تسحب من توزيع Y وهي مستقلة عن العينة الأساسية.

وينشأ عن ذلك:

$$Y_f = \hat{Y}_f + u_f$$

حيث ان u_f خطأ عشوائي جديد بالافتراضات السابقة نفسها وهو خطأ التنبؤ للعينة المختارة.

والذي متوسطه = صفر $(e_f = Y_f - \hat{Y}_f)$

ويتوزع توزيعاً طبيعياً

اما تباينه:

$$\operatorname{var}(e_f) = \operatorname{var}(Y_f) + \operatorname{var}(\hat{Y}_f)$$

$$= \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(X_f - \bar{X})^2}{S_{XX}} \right]$$

$$= \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{S_{XX}} \right] \quad \cdots \quad (10-3)$$

وهو اكبر من تباين متوسط الاستجابة. ⟸ وباختصار : $e_f \sim N(0, \operatorname{var}(e_f))$

وبذلك فان:

$$\frac{e_f}{\sigma \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum x^2}}} \sim N(0,1)$$

وحيث ان σ غير معلومة يتم تقديرها على وفق:

$$\hat{\sigma} = \sqrt{\frac{\sum e^2}{n-2}}$$

وبذلك:

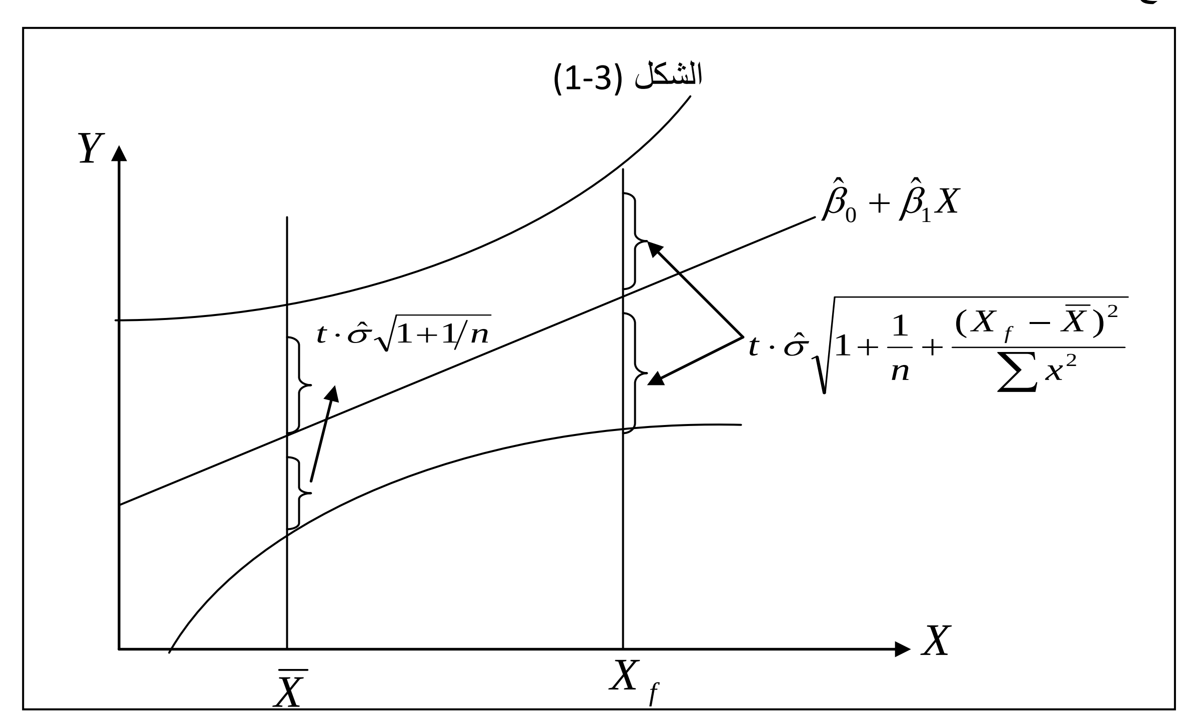
$$\frac{Y_f - \hat{Y}_f}{\sigma\sqrt{1+\frac{1}{n}+\frac{(X_f-\overline{X})^2}{\sum x^2}}} \sim t_{(n-2)} \quad \cdots \quad (11\text{-}3)$$

وهكذا فان فترة ثقة باحتمال 100(1-α)% للمشاهدات الجديدة Y_f عند X_f هي:

$$\hat{Y}_f - \hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_f-\overline{X})^2}{\sum x^2}}\cdot t_{c(n-2,\frac{\alpha}{2})} \le \ Y_f \ \le \hat{Y}_f + \hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_f-\overline{X})^2}{\sum x^2}}\cdot t_{c(n-2,\frac{\alpha}{2})} \quad \ldots (12\text{-}3)$$

وفترة الثقة للقيمة التنبؤية الجديدة تكون أوسع من فترة الثقة لمتوسط الاستجابة، هذا فضلاً عن ان فترة الثقة تزداد مع ابتعاد القيمة التنبؤية لـ $X(X_f)$ عن متوسط قيم $(\overline{X})$ X لكل من متوسط الاستجابة أو للقيمة التنبؤية الجديدة.

ويمكن توضيح ذلك من خلال الشكل (1-3)



<u>مثال(3-7):</u> بالاعتماد على معلومات المثال(3-1) لفرد في الحرفة نفسها خبرته(سنوات خدمته) (25) سنة فان فترة تنبؤية باحتمال 95% لأجره المتوقع يحسب على وفق الآتي:

$$\hat{Y}_{25} = 29.71 + 1.121(25)$$
$$= 57.735$$

$$\text{var}(\hat{Y}_{25}) = 70.67(1+\frac{1}{16}+\frac{(25-15.3125)^2}{1539.44}) = 70.67(1.063+0.061)$$
$$= 79.433$$
$$\hat{\sigma} = 8.913$$
$$57.735 - 8.913(2.145) \le Y_{25} \le 57.735 + 8.913(2.145)$$
$$38.617 \le Y_{25} \le 76.853$$

لذا فان فرداً بالحرفة نفسها وبخبرة (25) سنة، يتوقع ان يكون أجرة بين 76.853 الف دينار و 38.617 الف دينار.

ويمكن استخدام مجال الثقة لاختبار فيما اذا كانت المشاهدات الجديدة (X_f, Y_f) متولدة من هيكل المولد نفسه لمشاهدات العينة ام لا.

فمثلاً النقطة (25,80) تكون قيمة Yخارج مجال الثقة وبالتالي نستنتج بان هذه النقطة لاتتولد من هيكل العينة نفسه.

<u>مثال (3-8):</u> مع توافر البيانات التالية :

$$\hat{Y}_i = 2.5 + 0.7X_i \qquad i = 1,2,...,15$$

$$Tss = 290 \quad , \quad \sum X^2 = 260 \quad , \quad \sum X = 30$$

احسب حدود الثقة باحتمال 99% للقيمة التنبؤية الجديدة عندما $X_f = 10$.

الحل:

$$\hat{Y}_{10} = 2.5 + 0.7(10) = 9.5$$

$$\text{var}(\hat{Y}_{10}) = \frac{\sum e^2}{n-2} \cdot \left[1 + \frac{1}{n} + \frac{(10-\bar{X})^2}{S_{XX}}\right] \quad , \quad \bar{X} = 2$$

$$S_{XX} = \sum X^2 - \frac{(\sum X)^2}{n} = 260 - 60 = 200$$

$$Ess = \hat{\beta}_1^2 S_{XX} = 98$$

$$Rss = Tss - Ess = 290 - 98 = 192$$

$$\hat{\sigma}^2 = \frac{192}{13} = 14.769$$

$$s.e(\hat{Y}_{10}) = \sqrt{14.769(1.067+0.32)} = \sqrt{14.769}\sqrt{(1.387)} = \sqrt{20.485} = 4.526$$

حدود الثقة باحتمال 99% للقيمة التنبؤية الجديدة عندما $X_f = 10$:

$$9.5 - 4.526(3.012) \le Y_{10} \le 9.5 + 4.526(3.012)$$
$$9.5 - 13.632 \le Y_{10} \le 9.5 + 13.632$$
$$-4.132 \le Y_{10} \le 23.132$$

يمكن ان نستنتج من فترة الثقة بان القيمة التنبؤية الجديدة غير معنوية لان قيمة الصفر توجد ضمن فترة الثقة.

(5-3)تحليل التباين لنموذج الانحدار الخطي البسيط ANOVA for simple Linear Regression Model.

في هذه الفقرة ستتم دراسة تحليل التباين. ولابد من التأكيد هنا الى ان اختبار معنوية المتغير المستقلX . $(\mathrm{H}_0 : \beta_1 = 0)$ والذي تمت دراسته في المبحث (3-1)، يمكن ايضاً معالجته باستخدام اطار تحليل التباين، والجدول (3-1) يوضح ذلك.

يتكون الجدول من عدة أعمدة، العمود الاول يتمثل بتقسيم التغيرات الاجمالية في المتغيرY الى تغيرات مشروحة من قبل المتغير X ومتغيرات غير مشروحة متروكة لحد الخطأ العشوائي كما تم توضيحه في الفقرة (2-4-8) .

والعمود الثاني يمثل مجموع المربعات لفقرات العمود الاول، ويرمز له ss .

والعمود الثالث يمثل درجات الحرية لفقرات العمود الاول. ويرمز له d.f وتتميز عناصر كل عمود بان حاصل جمعها يمثل المجموع الكلي.

والعمود الرابع يتمثل بمتوسط المربعات (MS) يتم الحصول على عناصره بقسمة عناصر العمود الثاني (مجموع المربعات) لكل سطر على عدد درجات الحرية المناظر له في العمود الثالث.

ولايفوتنا التأكيد على ان التفسير البديهي لمفهوم درجات الحرية (d.f) هو عدد القيم التي يمكن وضعها بشكل كيفي (اعتباطي). فمثلاً لقيم y يمكن وضع (n-1) من قيمه كما نشاء. وتبقى المشاهدة الاخيرة (n) لابد من تحديدها تحت شرط $\Sigma y = 0$ وكذلك بالنسبة لقيم الاخطاء (e) يمكن تحديد (n-2) من قيمها كيفما اتفق. حيث ان تقدير المربعات الصغرى يضع قيدين على (e) وهما $\Sigma e = 0$ و $\Sigma Xe = 0$.

وبذلك تبقى درجة حرية واحدة مرتبطة بمجموع المربعات المشروحة والتي تعتمد على معلمة β_1.

أما العمود الاخير في الجدول فيتمثل بقيمة F والمتمثلة بقسمة متوسط المربعات المشروحة على متوسط المربعات غير المشروحة

$$F = \frac{MES}{MRS} = \frac{ESS/1}{RSS/(n-2)} \quad \ldots \quad (13\text{-}3)$$

وبذلك فان أثر المتغير X يمكن الاستدلال حوله واختيار أهميته لمعادلة الانحدار من خلال مقارنة القيمة العملية لـ F مع القيمة النظرية (الجدولية) والتي تستخرج من جداول F الخاصة بدرجات حرية(1) للبسط و (n-2) للمقام ومستوى دلالة 5% .

فاذا كانت:

$$F = \frac{ESS/1}{RSS/(n-2)} \rangle F_{(1,n-2,0.95)}$$

فيكون القرار برفض فرضية العدم: $(H_0 : \beta_1 = 0)$ عند مستوى دلالة 5% .

ان العلاقة (13-3) يتم الحصول عليها على وفق أساسيات الإحصاء الرياضي كآلاتي:

معلوم من مباحث الفصل الثاني ان:

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{S_{XX}}} \sim N(0,1)$$

$$\Rightarrow \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2/S_{XX}} \sim \chi^2{}_{(1)}$$

وكذلك: $\frac{\Sigma e^2}{\sigma^2} \sim \chi^2{}_{(n-2)}$

$$\Rightarrow F = \frac{(\hat{\beta}_1 - \beta_1)^2 \Sigma x^2}{\Sigma e^2/(n-2)} \sim F_{(1,n-2)}$$

حيث ان F هي النسبة لمتغيرين مستقلين كل منهما ذو توزيع χ^2 مقسوم على عدد درجات الحرية المناسبة. وتحت فرضية العدم: $H_0 : \beta_1 = 0$

$$F = \frac{\hat{\beta}_1{}^2 \Sigma x^2}{\Sigma e^2/(n-2)} \quad \ldots \quad (14-3)$$

وباعتماد تجزئة مجموع المربعات الكلية فان العلاقة (14-3) ستؤول الى:

$$F = \frac{ESS/1}{RSS/(n-2)}$$

وهي بالضبط العلاقة (13-3)

جدول(1-3)

جدول تحليل التباين للانحدار البسيط

- ANOVA-

Source of variation مصدر المتغيرات	Sum of squares مجموع المربعات (SS)	Degree of freedom درجات الحرية (d.f)	Mean squares MSS متوسط مجموع المربعات
(x) : الانحرافات الموضحة	$ESS = \Sigma\hat{y}^2 = \hat{\beta}_1^2\Sigma x^2 = \hat{\beta}_1\Sigma xy$	١	$ESS/1$
(البواقي): الانحرافات غير الموضحة	$RSS = \Sigma e^2$	n-2	$RSS/n-2$
الإجمالي: المجموع الكلي	$TSS = \Sigma y^2$	n-1	

مثال (9-3) : بالرجوع الى معلومات المثال (1-3) فان اختبار أهمية المتغير التوضيحي (عدد سنوات الخدمة) في تحديد الاجر، يمكن اختباره على وفق جدول تحليل التباين كالاتي:

vs $\mathrm{H}_0 : \beta_1 = 0$ $\mathrm{H}_1 : \beta_1 \neq 0$

تحسب قيمة F للعينة:

$$ESS = 1934.42$$
$$RSS = 989.33$$
$$n-2 = 14$$
$$\Rightarrow F = \frac{1934.42/1}{989.33/14} = \frac{1934.42}{70.67} = 27.37$$

وبمقارنتها مع القيمة الجدولية: $F_{(1,14,0.95)} = 4.60$

قيمة F للعينة اكبر من القيمة الجدولية ⟸ نرفض فرضية العدم، أي ان المعلمة β_1 معنوية أي ان المتغير X مهم في تحديد التغيرات في Y .

والنتيجة متماثلة مع مثال(1-3) الذي تم استخدام نسبة t للاختبار.

ويمكن وضع النتائج في الجدول (2-3)

جدول(3-2)

جدول تحليل التباين للمثال (3-9)

S.o.v	SS	d.f	MSS
X	$ESS = 1934.42$	١	1934.42
Error	$RSS = 989.33$	14	70.67
Total	$TSS = 2923.75$	16	

ملاحظة: لابد من التنويه هنا ان استخدام جدول تحليل التباين يصلح لاختبار الفرضية $(H_0 : \beta_1 = 0)$ حصراً. اما اختبار الفرضيات حول معنوية المقطع الصادي أو اختبار الفرضية ذات الطرف الواحد أو اختبار الفرضيات بان β_1 تساوي قيمة معينة فلا يصح إلا باستخدام الاحصاءة t .

(3-6)<u>اختبار حسن التوفيق</u> <u>Goodness of Fit</u>

ويسمى ايضاً مقياس جودة التقدير أو معامل التحديد (Coefficient of determination) وهو مقياس يقيس تشتت المشاهدات حول خط الانحدار.

فهو يمثل الأهمية النسبية للتغيرات المشروحة من قبل معادلة التقدير. ويرمز له R^2:

$$\left.\begin{aligned} R^2 &= \frac{ESS}{TSS} \\ &= 1 - \frac{RSS}{TSS} \end{aligned}\right\} \quad \cdots \qquad (15-3)$$

مثال(3-10) : من معادلة الانحدار:

$$\hat{Y}_i = 2.3 + 0.7 X_i \qquad i = 1,2,\ldots,20$$

$$\Sigma(Y - \bar{Y})^2 = 135 \quad , \quad \Sigma X^2 = 260 \quad , \quad \Sigma X = 30$$

$$R^2 = \frac{ESS}{\Sigma(Y - \bar{Y})^2}$$

$$ESS = \hat{\beta}_1^{\,2} \Sigma x^2 = (0.7)^2 \left[260 - \frac{(30)^2}{20} \right] = (0.49)(215)$$

$$\therefore R^2 = \frac{105.35}{135} = 0.78$$

78% من التغيرات الإجمالية في المتغير المعتمد Y تم تفسيرها بواسطة المتغير التوضيحي X .

22% فقط من هذه التغيرات الإجمالية تركت لحد الخطأ.

فهو بذلك يقيس تشتت المشاهدات حول خط الانحدار.

فكلما كانت المشاهدات قريبة من الخط المقدر يعطي دلالة على ان التقدير جيد. وعليه فان R^2 تقيس مقدار التغيرات في Y التي تم توضيحها عن طريق التغيرات في X .

اما قيمة هذا المؤشر فتمتلك الصفتين:

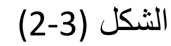
١- قيمته غير سالبة لانه نسبة بين مجموع المربعات المشروحة الى الكلية.

٢- حدود قيمته: $0 \leq R^2 \leq 1$

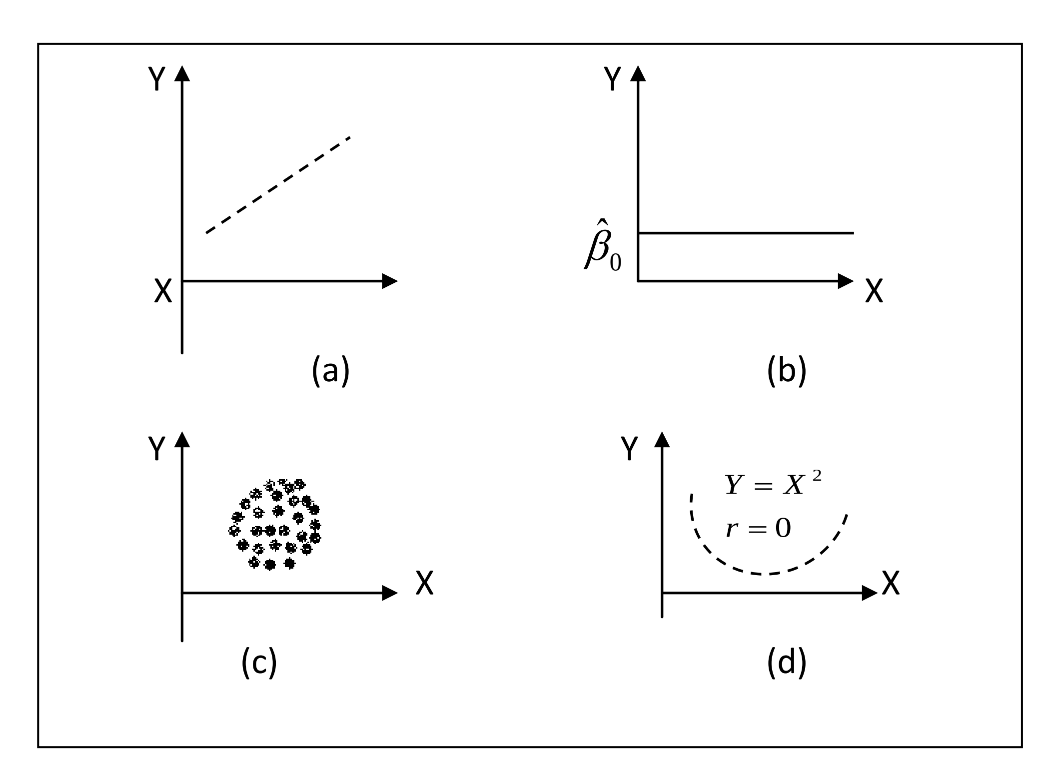
فاذا كانت القيمة تساوي واحد يعني ان جميع مشاهدات العينة تقع على معادلة الخط المقدر، أي ان العلاقة بين Y و X تامة (perfect) ، أي ان $(\hat{Y}_i = Y_i \quad \forall \quad i)$ كما في (a) من الشكل (2-3).

وعلى النقيض من ذلك اذا كانت $R^2=0$ فهذا يعني لاتوجد علاقة بين Y و X (بمعنى $\hat{\beta}_1 = 0$) وبذلك $\hat{Y}_i = \hat{\beta}_0 = \bar{Y}$. بمعنى ان افضل تنبؤ لقيم Y يكون ببساطة ممثلة بقيمة المتوسط وفي هذه الحالة فان الخط المستقيم المقدر يكون خطاً موازياً للمحور X كما في الشكل (b) .

ولابد من التأكيد بان R^2 قد يكون مقياساً غير جيد للدلالة على ملاءمة النموذج للبيانات كما في (c) و (d) من الشكل (2-3) .



الشكل (2-3)



فالشكلان (c) و (d) يوضحان إذا كانت $R^2 = 0$ فهذا لايعني ان المتغيرين مستقلان.

مثال (3-11) : بيانات الجدول (2-1) للمثال (2-1)

معامل التحديد:

$$R^2 = \frac{ESS}{TSS} = \frac{221966.242}{240407.309}$$
$$= 0.92329$$

هذا يعني ان 92% من التغيرات في المتغير Y تم تحديدها بواسطة التغيرات في المتغير X ، كما ان (1-92.32976) أي 7.67% من التغيرات لم تتمكن معادلة الانحدار توضيحها بل تركت لحد الخطأ.

كما يمكن حساب معامل التحديد من أحد القوانين التالية:

$$\left.\begin{aligned} R^2 &= \frac{ESS}{TSS} \\ &= \frac{\Sigma\hat{y}^2}{\Sigma y^2} \\ &= \frac{\hat{\beta}_1{}^2\Sigma x^2}{\Sigma y^2} = \hat{\beta}_1{}^2\left(\frac{\Sigma x^2}{\Sigma y^2}\right) \\ &= \frac{\hat{\beta}_1\Sigma xy}{\Sigma y^2} = \hat{\beta}_1\left(\frac{\Sigma xy}{\Sigma y^2}\right) \end{aligned}\right\} \quad \cdots \quad (16-3)$$

(3-7)معامل الارتباط البسيط Simple Correlation Coefficient

ان تحليل الارتباط مختلف تماماً عن مفهوم الانحدار. فتحليل الارتباط هو مقياس لدرجة الترابط الخطي بين متغيرين ويرمز لمعامل ارتباط المجتمع بالرمز (ρ) اما في حالة العينة فيرمز له (r) والذي يعد تقديراً لمعامل ارتباط المجتمع (ρ) ويسمى معامل ارتباط بيرسن (Pearson). وبادئ ذي بدء لابد من توضيح الفرق المفاهيمي بين تحليل الانحدار وتحليل الارتباط. ففي تحليل الانحدار يوجد عدم تماثل في طريقة معاملة المتغير المعتمد والمتغير المستقل. فالمتغير المعتمد يفترض ان يكون عشوائيا ويمتلك توزيعاً احتمالياً. في حين المتغير المستقل يفترض ان تكون قيمه ثابتة للعينة المختارة. وعلى الجانب الآخر في تحليل الارتباط يتم معاملة المتغيرين بشكل متماثل. فلا يوجد اختلاف بين المتغير المعتمد والمتغير المستقل فكلاهما يفترض ان يكون عشوائياً.

ان الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين.

ويعرف معامل الارتباط بين متغيرين بأنه يتمثل في نسبة التغاير بين المتغيرين الى حاصل ضرب انحرافهما المعياريين.

$$r=\hat{\rho}_{xy}=\frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2(Y-\bar{Y})^2}}=\frac{\Sigma XY-\Sigma X\Sigma Y/n}{\sqrt{\Sigma X^2-\frac{(\Sigma X)^2}{n}}\cdot\sqrt{\Sigma Y^2-\frac{(\Sigma Y)^2}{n}}}=\frac{\Sigma xy}{\sqrt{\Sigma x^2\Sigma y^2}}\quad ...\ (17\text{-}3)$$

علاقة معامل الارتباط مع معلمة الانحدار $\hat{\beta}_1$:

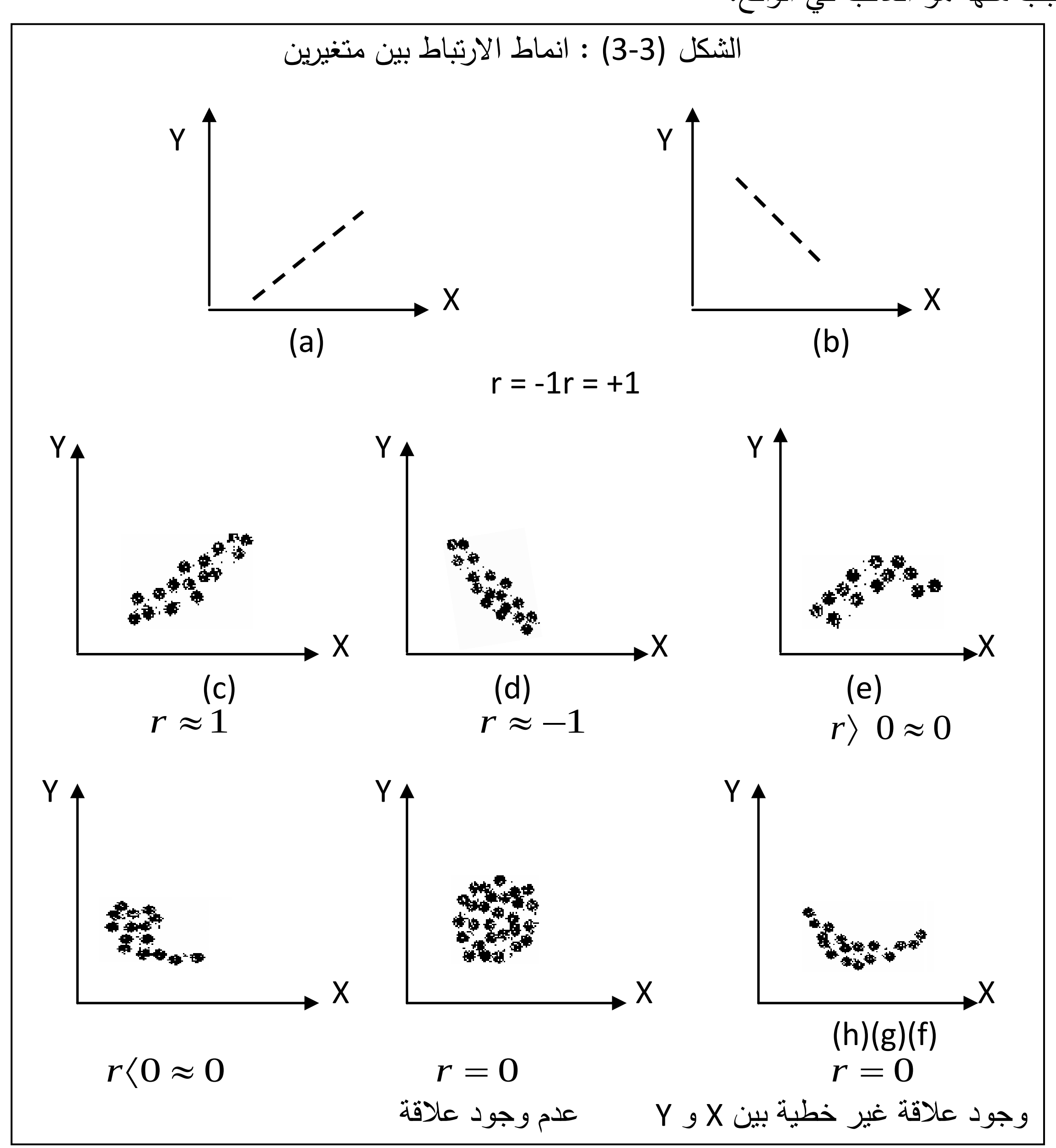
$$r=\frac{\Sigma xy}{\Sigma x^2}\cdot\frac{\sqrt{\Sigma x^2}}{\sqrt{\Sigma y^2}}=\hat{\beta}_1\sqrt{\frac{S_{XX}}{S_{YY}}}\quad ...\quad (18\text{-}3)$$

خواص معامل الارتباط البسيط:

١– اشارة معامل الارتباط تعتمد على اشارة التغاير بين المتغيرين. فقد تكون موجبة (طردية) أو سالبة (عكسية) كما في (a) و (b) من الشكل (3-3).

والموجب منها هو الغالب في الواقع.

الشكل (3-3) : انماط الارتباط بين متغيرين



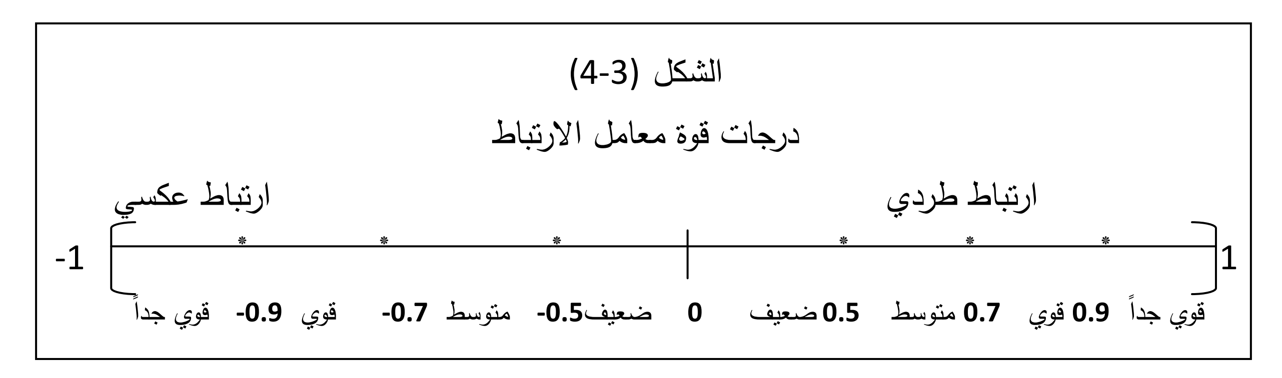
٢- متماثلة أي($r_{XY} = r_{YX}$) وخالية من الوحدات ومستقلة عن نقطة الأصل. اذا كانت $X^* = aX + c$ و $Y^* = bY + d$ ، b > 0 ، a > 0 ، c ، d ثوابت، فان معامل ارتباط X و Y هو نفس معامل ارتباط X^* و Y^* :

$$r_{XY} = r_{X^*Y^*}$$

3- هو مقياس لدرجة الترابط الخطي فقط وبذلك قد يكون معامل الارتباط (صفر) بين متغيرين وهو لايدل على انعدام الارتباط بينهما فقد تكون علاقة غير خطية عالية بين المتغيرين كما في الشكل(h).

٤ - بالرغم من ان المقياس يوضح درجة الترابط الخطي بين متغيرين ولكنه في الوقت نفسه لا يبرر وجود علاقة سببية بين المتغيرين فربما ينشأ الترابط بسبب وجود متغير ثالث يظهر ذلك الترابط.

5- قوة العلاقة بين متغيرين يمكن تحديدها من حيث درجة قربها أو بعدها عن (1 ±)حيث ان المدى لمعامل الارتباط البسيط هو: (1 > r > -1) ويمكن تصنيف درجات القوة كما في الشكل (4-3)



الشكل (4-3)
درجات قوة معامل الارتباط

<u>مثال (12-3):</u> افترض المساحة المزروعة بالأعلاف الخضراء (ألف هكتار) (X) واجمالي انتاج اللحوم (Y) (ألف طن) للسنوات 1999-2006.

السنة	1999	2000	2001	2002	2003	٢٠٠٤	2005	2006	المجاميع
المساحة X	305	313	297	289	233	214	240	217	2108
الإنتاج Y	592	603	662	607	635	699	719	747	5264

المطلوب: حساب معامل الارتباط بين المساحة المزروعة وكمية إنتاج اللحوم. ووضح دلالته الإحصائية.

الحل: بتطبيق العلاقة (3-17) على وفق الاتي:

اولاً: المتغيرات كانحرافات عن المتوسط.

1- حساب المتوسط الحسابي لكل من X و Y:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{2108}{8} = 263.5$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{5264}{8} = 658$$

2- نحسب X و Y كانحرافات عن متوسطاتها:

$$y = Y - \bar{Y} \qquad , \qquad x = X - \bar{X}$$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
305	592	41.5	-66	-2739	1722.25	4356
313	603	49.5	-55	-2722.5	2450.25	3025
297	662	33.5	4	134	1122.25	16
289	607	25.5	-51	-1300.5	650.25	2601
233	635	-30.5	-23	701.5	930.25	529
214	699	-49.5	41	-2029.5	2450.25	1681
240	719	-23.5	61	-1433.5	552.25	3721
217	747	-46.5	89	-4138.5	2162.25	7921

3- نحسب المجاميع:

$$\Sigma xy = S_{xy} = -13528 \ , \ \Sigma x^2 = S_{xx} = 12040 \ , \ \Sigma y^2 = S_{yy} = 23850$$

4- نطبق العلاقة (3-17)

$$r = \frac{-13528}{\sqrt{12040}\sqrt{23850}}$$

$$r = \frac{-13528}{(109.727)(154.434)} = \frac{-13528}{16945.619} = -0.798$$

الدلالة الإحصائية: يوجد ارتباط عكسي قوي بين المساحة المزروعة وكمية انتاج اللحوم. فزيادة المساحات المزروعة تسهم بانخفاض المساحات المحددة للرعي وهذه بدورها تؤدي الى انخفاض الثروة الحيوانية أي انخفاض انتاج اللحوم.

ثانياً: باستخدام المتغيرات الاصلية:

1- نحسب المجاميع:

X	Y	XY	X^2	Y^2
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244
289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009
2108	5264	1373536	567498	3487562

$$r = \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{567498 - \frac{(2108)^2}{8}}\sqrt{558009 - \frac{(5264)^2}{8}}}$$

$$= \frac{-13528}{\sqrt{12040}\sqrt{23850}} = \frac{-13528}{16945.619}$$

$$= -0.798$$

وهي النتيجة السابقة نفسها.

اختبار معنوية الارتباط البسيط:

معامل الارتباط بين المتغيرين X و Y للمجتمع قيمته مجهولة ρ_{XY} وتم تقدير معامل الارتباط للعينة $(\hat{\rho}_{XY} = r_{XY})$ ، وللتحقق من معنوية المعلمة المقدرة (r_{XY}) يتم إتباع التالي:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$t^* = \frac{r - 0}{\sqrt{\frac{1 - r^2}{n - 2}}} \sim t_{(n-2, \frac{\alpha}{2})} \qquad . . . \qquad (19\text{-}3)$$

مثال(3-13): اختبر الفرضية التالية $H_0 : \rho = 0 \quad vs \quad H_1 : \rho \neq 0$

بالاعتماد على معلومات المثال (3-12)

وبتطبيق العلاقة (3-19):

$$t^* = \left| \frac{-0.798}{\sqrt{\frac{1-(-0.798)^2}{8-2}}} \right| = \left| \frac{-0.798}{\sqrt{\frac{1-637}{6}}} \right| = \left| \frac{-0.798}{\sqrt{0.06}} \right|$$

$$t^* = \left| \frac{-0.798}{0.246} \right| = 3.244$$

$$t_{c(6,0.025)} = 2.447$$

وبمقارنة قيمة t^* المحسوبة (3.244) مع القيمة الجدولية t_c(2.447) يتضح ان:

$$t^* \rangle \ t_c$$

فيكون القرار بمعنوية معامل الارتباط احصائياً.

ان هذا الاختبار هو مطابق تماماً لاختبار: $H_0 : \beta_1 = 0$

علاقة معامل الارتباط البسيط ومعامل التحديد في نموذج الانحدار البسيط:

يمكن استخدام معامل الارتباط البسيط (r) في قياس قوة العلاقة في معادلة الانحدار البسيط.

فمعامل الارتباط البسيط بين Y و $\hat{Y}$ هو $(\rho_{Y\hat{Y}})$ يمكن استخدامه لقياس مدى اقتراب Y و $\hat{Y}$.

وهو بذلك يعد مقياساً لقدرة نموذج الانحدار على تفسير قيم Y.

ويمكن برهنة ذلك على وفق الآتي:

$$r_{XY} = \sqrt{R^2}$$

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$= \frac{\Sigma xy}{\Sigma x^2} \cdot \frac{\Sigma x^2}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{\hat{\beta}_1 \sqrt{\Sigma x^2}}{\sqrt{\Sigma y^2}} = \left[\frac{\hat{\beta}_1{}^2 \Sigma x^2}{\Sigma y^2}\right]^{1/2}$$

$$= \left[\frac{ESS}{TSS}\right]^{1/2} = \sqrt{R^2}$$

كما يمكن برهنة $(r_{y\hat{y}} = \sqrt{R^2})$ على وفق الآتي:

$$r_{Y\hat{Y}} = \frac{\Sigma(Y-\overline{Y})(\hat{Y}-\overline{\hat{Y}})}{\sqrt{\Sigma(Y-\overline{Y})^2 \Sigma(Y-\overline{\hat{Y}})^2}}$$

ولكن:

$$\overline{Y} = \frac{\Sigma Y}{n} = \frac{\Sigma(\hat{Y}+e)}{n} = \overline{\hat{Y}}$$

البسط:

$$\Sigma(Y-\overline{Y})(\hat{Y}-\overline{Y}) = \Sigma\left[(\hat{Y}+e)-\overline{Y}\right]\ \left[\hat{Y}-\overline{Y}\right]$$

$$= \Sigma\left[(\hat{Y}-\overline{Y})+e\right]\ \left[\hat{Y}-\overline{Y}\right]$$

$$= \Sigma(\hat{Y}-\overline{Y})(\hat{Y}-\overline{Y}) + \Sigma e(\hat{Y}-\overline{Y})$$

$$= \Sigma(\hat{Y}-\overline{Y})^2 + \Sigma e\hat{Y} - \overline{Y}\Sigma e$$

لان:

$$\Sigma e\hat{Y} = \Sigma e(\hat{\beta}_0 + \hat{\beta}_1 X) = \hat{\beta}_0 \Sigma e + \hat{\beta}_1 \Sigma eX$$

$$= 0$$

$$\Sigma(Y-\overline{Y})(\hat{Y}-\overline{Y}) = \Sigma(\hat{Y}-\overline{Y})^2 = ESS$$

المقام:

$$\sqrt{\Sigma(Y-\overline{Y})^2 \Sigma(\hat{Y}-\overline{Y})^2} = \sqrt{(TSS)(ESS)}$$

$$r_{Y\hat{Y}} = \sqrt{\frac{ESS}{TSS}} = \sqrt{R^2} \qquad . \; . \; . \qquad (20-3)$$

بعبارة اخرى فان مربع معامل الارتباط بين متغيرين (X_i) و (X_j) يقيس جزء التغيرات المشتركة بين هذين المتغيرين فاذا كان (r_{XY} = .56)(r^2=0.31) أي ان 31% من التباين يكون مشتركاً بين المتغير X و Y .

جدول تحليل التباين بدلالة معامل التحديد:

بالاعتماد على العلاقة (3-16) يمكن إعادة كتابة الاحصاءة Fعلى وفق الآتي:

$$F = \frac{ESS}{RSS/(n-2)} = \frac{R^2 \;\; TSS}{(1-R^2) \;\; TSS/(n-2)} = \frac{R^2}{(1-R^2)/(n-2)}$$

حيث:

$$R^2 = \frac{ESS}{TSS} \qquad \Rightarrow \qquad ESS = R^2 \; TSS$$

$$= R^2 \Sigma(Y-\bar{Y})^2$$

و

$$R^2 = 1 - \frac{RSS}{TSS} \qquad \Rightarrow \qquad RSS = (1-R^2) \; TSS$$

$$= (1-R^2) \;\; \Sigma(Y-\bar{Y})^2$$

وبذلك فان جدول تحليل التباين يمكن إعادة صياغته كآلاتي:

جدول(3-2)

جدول تحليل التباين ANOVA Table

S.o.v	SS	d.f	MSS
X	$R^2 \; \Sigma(Y-\bar{Y})^2$	١	$R^2\Sigma(Y-\bar{Y})^2$
البواقي	$-R)^2 \;\; \Sigma(Y-\bar{Y})^2$	n-2	$(1-R^2) \;\; \Sigma(Y-\bar{Y})^2/(n-2)$
Total	$\Sigma(Y-\bar{Y})^2$	n-1	

مثال(3-14): استخدم بيانات المثال (3-10) في تحديد جدول تحليل التباين.

ملخص البيانات: $n = 20$, $R^2 = 0.78$, $\Sigma(Y - \bar{Y})^2 = 135$

فجدول تحليل التباين:

S.o.v	SS	d.f	MSS	F^*
X	(0.78)(135)	١	105.3	63.8
Error	(0.22)(135)	18	1.65	
Total	135	19		

كما ويمكن معرفة معنوية العلاقة الخطية المفترضة بين المتغير المعتمد والمتغير المستقل من خلال مقارنة قيمة F النظرية (الجدولية) بدرجات حرية (1) البسط و (18) للمقام وباستخدام مستوى دلالة معين.

مثال (3-15): اكمل جدول تحليل التباين لاختبار خطية العلاقة بين X و Y اذا توفرت لديك المعلومات:

$$n = 50 \ , \ r_{XY} = 0.63 \ , \ TSS = 500$$

الحل:

$$R^2 = r^2$$
$$= 0.397$$

S.o.v	SS	d.f	MSS	F^*
X	$(500)(0.63)^2$	١	198.5	31.61
Error	(500)(1-0.397)	48	6.28	
Total	500	49		

لاختبار معنوية العلاقة الخطية:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

$$F^* = \frac{M(ESS)}{M(RSS)} = \frac{198.5}{6.28} = 31.61$$

وبمقارنتها بالقيمة الجدولية لمستوى الدلالة 5% و 1% على التوالي:

$$F_{(1,48,0.95)} = 4.04$$

$$F_{(1,48,0.99)} = 7.19$$

لذا نرفض H_0 أي ان العلاقة الخطية معنوية.

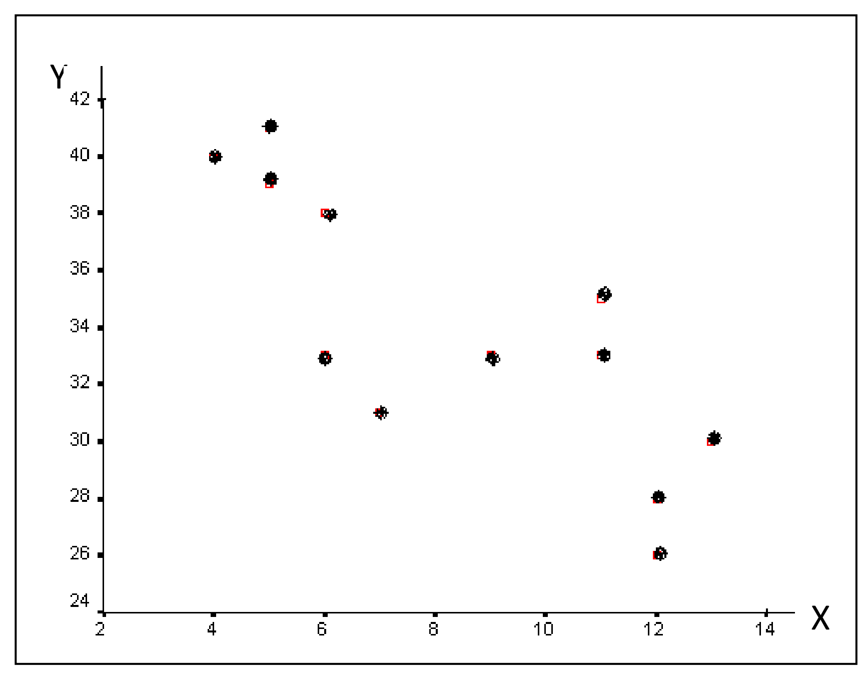
<u>مثال (16-3) تطبيقي شامل:</u>

في دراسة العلاقة بين عيوب المنتج (Y) مقاسة بعدد العيوب الموجودة في المشاهدة وخبرة العامل مقاسة بعدد سنوات الخدمة (X) ، لعينة من (15) مشاهدة.

رقم المشاهدة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
عيوب المنتج (Y)	39	35	40	28	35	30	33	33	26	40	41	33	31	33	38
الخبرة (X)	5	11	4	12	11	13	9	11	12	4	5	6	7	9	6

$$\Sigma X = 125 \; , \; \Sigma Y = 515 \; , \; \Sigma X^2 = 1185 \; , \; \Sigma XY = 4128 \; , \; \Sigma Y^2 = 17973$$

- رسم الانتشار:



رسم الانتشار لبيانات المثال (16-3) باستخدام برنامج SPSS

من رسم الانتشار يتضح ان العلاقة الخطية هي المرشحة

- $$\hat{\beta}_1 = \frac{\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i / n}{\Sigma X_i^2 - (\Sigma X_i)^2 / n} = \frac{4128 - \frac{(125)(515)}{15}}{1185 - \frac{(125)^2}{15}} = \frac{-163.67}{143.3} = -1.142$$

- $$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 34.34 + 1.142(8.34) = 43.85$$

- ∴ معادلة التقدير: $$\hat{Y}_i = 43.85 - 1.142 X_i$$

زيادة سنوات الخبرة سنة واحدة تؤدي الى انخفاض عيوب المنتج بمقدار 1≈1.142 .

- <u>اختبار معنوية المعلمات:</u>

المقطع الصادي:

vs $H_0 : \beta_0 = 0 \;\; H_1 : \beta_0 \neq 0$

$$\text{var}(\hat{\beta}_0) = \hat{\sigma}^2 \frac{\Sigma X^2}{n S_{xx}} \quad ; \quad S_{xx} = \Sigma X^2 - \frac{(\Sigma X)^2}{n} = 143.3$$

$$\hat{\sigma}^2 = \frac{\Sigma e^2}{n-2} \quad ; \quad \Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 291.33 \quad ;$$

$$ESS = \hat{\beta}_1^2 S_{xx} = (1.142)^2 (143.3) = 186.88$$

$$RSS = TSS - ESS = 104.45$$

$$\hat{\sigma}^2 = \frac{104.45}{13} = 8.035$$

$$\text{var}(\hat{\beta}_0) = 8.035 \cdot \frac{1185}{15(143.3)} = 8.035 \cdot \frac{1185}{2149.5} = 4.43$$

$$; t_{\hat{\beta}_0} = \frac{43.85}{\sqrt{4.43}} = \frac{43.85}{2.105} = 20.83 \quad t_{c(13,0.025)} = 2.16$$

$t^*_{\hat{\beta}_0} \rangle t_c \Leftarrow$ نرفض فرضية العدم. المقطع الصادي معنوي إحصائياً.

- <u>الميل $\hat{\beta}_1$:</u> $H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$

$$\text{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}} = \frac{8.035}{143.3} = 0.056$$

$$\left|t^*_{\hat{\beta}_1}\right| = \left|\frac{-1.142}{0.237}\right| = 4.82$$

$$t_{c(13,0.025)} = 2.16$$

$t^*_{\hat{\beta}_1} \rangle t_c \Leftarrow$ نرفض H_0 . أي ان المتغير المستقل مهم لتحديد التغيرات في Y .

- مجال الثقة لـ β_0 : $43.85 \pm 2.105(2.16) \equiv 43.85 \pm 4.547$

أي: $(39.303 \quad , \quad 48.396)$

- مجال الثقة لـ β_1 : $-1.142 \pm 0.237(2.16) \equiv -1.142 \pm 0.512$

$(-1.654 \quad , \quad -0.63)$

- معامل التحديد:

$$R^2 = \frac{ESS}{TSS} = \frac{186.88}{291.33} = 0.641$$

64 % من التغيرات في Y (عيوب المنتج) يتم توضيحها من قبل معادلة الانحدار باستخدام خبرة العامل.

ANOVA Table

S.o.v	SS	d.f	MSS	F*
X	186.88	١	186.88	23.258
Error	104.45	13	8.035	
Total	291.33	14		

- مجال الثقة لمتوسط الاستجابة:

متوسط الاستجابة عندما تكون سنوات خدمة العامل = 10 سنوات

$$\hat{Y}_{10} = 43.85 - 1.142(10) = 32.43$$

$$\text{var}(\hat{Y}_{10}) = \hat{\sigma}^2\left[\frac{1}{n} + \frac{(10-\overline{X})^2}{\Sigma x^2}\right]$$

$$= 8.035\left[\frac{1}{15} + \frac{(2.89)}{143.3}\right] = 8.035(0.07 + 0.02) = 0.723$$

$$s.e(\hat{Y}_{10}) = 0.85$$

حدود متوسط الاستجابة عندما $X_f = 10$ هي

$$(32.43 \pm 0.85(2.16)) \equiv (32.43 \pm 1.84)$$

$$(30.59 \quad , \quad 34.27)$$

المشاهدة الجديدة: (10,30) لاتتولد من هيكل العينة نفسه، لان قيمة Y تكون خارج حدود مجال الثقة 34.266 و 30.59 .

– مجال الثقة للقيمة التنبؤية الجديدة عندما خدمة العامل = 10 سنوات

$$\hat{Y}_{10} = 32.43$$

$$\text{var}(Y_{10}) = \hat{\sigma}^2 + \text{var}(\hat{Y}_{10}) = 8.035 + 0.723 = 8.758$$

$$s.e(Y_{10}) = 2.96$$

حدود الثقة عندما $X_f = 10$ هي:

$$(32.43 \pm 2.96(2.16)) \equiv (32.45 \pm 6.39)$$

$$(26.04 \quad , \quad 38.824)$$

نفرض ان العيوب التي ينتجها عامل له خدمة 10(سنوات) = 12

أي ان المشاهدات الجديدة (10,12) لاتنتمي إلى هيكل العينة موضع الدراسة. في حين ان المشاهدات الجديدة (10,20) تنتمي إلى هيكل العينة.

في جميع الامثلة السابقة تم استخدام الطرائق الحسابية الاعتيادية للحل. ويمكن استخدام طريقة البرنامج الجاهز SPSS لجميع هذه الحسابات. وسيتم توضيح عرض نتائج SPSS.

مثال (3-17) لدراسة جودة منتج معين تم اختيار عينة من خمس عشرة مشاهدة وتم قياس عدد العيوب في كل منتج من المشاهدات (Y) واعتمد على خبرة العامل بقياسها بعدد سنوات الخدمة (X_1) ، والبيانات معروضة في الجدول (3-3)

الجدول (3-3)

عدد العيوب (Y) وسنوات الخدمة (X_1) لعينة من (15) مشاهدة سحبت عشوائياً من مصنع معين.

Y	39	٣٥	٤٠	28	35	30	33	33	26	40	41	33	٣١	33	38
X_1	5	11	4	12	11	13	9	11	12	4	5	6	٧	٩	٦

المصدر : http // samehar. word press.com

ومن أجل الاجابة عن الفقرات التالية:

١. تقدير معادلة الانحدار.

٢. اختبار معنوية المعلمات المقدرة.

٣. اختبار معنوية النموذج.

٤. حساب معامل التحديد.

يتم استخدام برنامج SPSS فتظهر نتائج التقدير وكما موضحة في الجدول(3-4):

جدول (3-4)

نتائج التقدير الخاصة بالمعاملات

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	β (1)	Std. Error (2)	Beta (3)	(4)	(5)
(Constant)	43.849	2. 041		20.837	.000
X_1	-1.142	.237	-.801	-4.823	.000

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

فالعمود الاول يمثل قيم معلمات النموذج المقدرة: $\hat{\beta}_0 = 43.849$ و $\hat{\beta}_1 = -1.142$.

والعمود الثاني يمثل قيم الخطأ المعياري للمعلمات بالتناظر $s.e(\hat{\beta}_0) = 2.041$ وكذلك $s.e(\hat{\beta}_1) = 0.237$

في حين القيم في العمود (4) تمثل قيم الاحصاءة t المحسوبة لكل معلمة والتي تمثل $t^*_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{s.e(\hat{\beta}_i)}$

فان قيمة t المحسوبة للمعلمة $\hat{\beta}_0$ هي (20.837) وكذلك قيمة t المحسوبة للمعلمة $\hat{\beta}_1$ هي (-4.823).

أما العمود (5) من الجدول فيشير الى قيم p وهو الاحتمال الذي تكون المعلمة المعنية معنوية. فاذا كانت قيمة p أقل من ٥%($p < 0.05$) فدليل على معنوية المعلمة المعنية باستخدام مستوى دلالة ٥%.

يمكن عرض النتائج بالصيغة التالية:

$$\hat{Y} = 43.849 - 1.142\,X_1$$
$$s.e: (2.104) \qquad (0.237)$$
$$t \quad :(20.837)^* \; (-4.823)^*$$

اذ تشير (*) الى ان المعلمة المعنية معنوية باستخدام ٥%.

ويمكن تفسير المعلمات المقدرة كالاتي:

$\hat{\beta}_0 = 43.849$ (1) تمثل أثر المتغيرات غير المضمنة في معادلة النموذج على عيوب المنتج وهي قيمة موجبة وكبيرة، وتدل على ان هناك متغيرات اخرى مهمة محذوفة من معادلة التقدير مثل المواد الاولية المستخدمة، او طرق التصنيع. . . الخ.

وتشير $\hat{\beta}_1 = -1.142$ الى ان زيادة سنة اضافية من سنوات الخبرة لدى العامل كفيلة بخفض العيوب في المنتج بمقدار (1.142).

(2) ان قيمة $s.e(\hat{\beta}_i)$ الانحراف المعياري للمعلمات المقدرة منخفضة وهي أقل من ($\frac{1}{2}$ قيمة المعلمة المقدرة) فهذا دليل على معنوية المعلمات والتي تؤكدها قيم t المحسوبة للمعلمات: $t^*_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{s.e(\hat{\beta}_i)}$ ، i = 1,2

$t^*_{\hat{\beta}_0} = 20.837$ وهي كبيرة مقارنة بالقيمة الجدولية باحتمال 5% وبدرجات حرية (n-2) وهذا ما تؤكده قيم p المناظرة لها وهي أقل من 5% ، أي ان معلمة المقطع الصادي معنوية باحتمال 5%.

وكما ان القيمة المطلقة للاحصاءة t الخاصة بمعلمة الانحدار $\left| t^*_{\hat{\beta}_1} \right| = 4.823$ هي الاخرى اكبر من قيمة الاحصاءة t الجدولية وهذا ما تؤكده قيم p المناظرة.

لذا نستنتج ان المعلمة $\hat{\beta}_1$ معنوية احصائيا باحتمال 5% . أي ان سنوات الخبرة تعد من المتغيرات المهمة التي تؤثر في عيوب المنتج حسب نتائج العينة المستخدمة.

(3) كما ان مخرجات البرنامج توفر جدول تحليل التباين والذي يساعد في اختبار معنوية النموذج ككل على وفق اختبار F وكالاتي:

جدول (3-5)

نتائج تحليل التباين

ANOVA

s.o.v	SS (1)	d.f (2)	MSS (3)	F (4)	Sig (5)
Regression	186.884	1	186.884	23.260	.000
Residual	104.449	13	8.035		
Total	291.333	14			

المصدر : نتائج برنامج SPSS بالاعتماد على بيانات العينة.

فالجدول (3-5) يوضح نتائج تحليل تباين الانحدار(ANOVA) الذي من خلاله يتم اختبار معنوية النموذج ككل، فالعمود (1) يمثل قيم مجموع مربعات التغيرات لكل فقرة من فقرات الجدول وهي الانحدار والبواقي والتغيرات الاجمالية، وهي على التوالي 186.884 ، 104.449 ، 291.333 .

أما القيم المعروضة في العمود(2) من الجدول فتشير الى درجات الحرية وعلى التوالي للانحدار= 1، للبواقي = 13 ، وللمجموع = n-1 = 14 .

والعمود(3) هو متوسطات القيم (MS) والتي تمثل مجموع المربعات لكل فقرة مقسومة على درجات الحرية المناسبة لها.

والعمود (4) يمثل قيم F الحسابية $(F = \frac{EMS}{RMS} = 23.260)$ والقيمة كبيرة مقارنة بالقيم الجدولية بدرجات حرية (1) للبسط و (13) للمقام وهذا ما تدعمه القيمة في العمود (5) ان (P = 0.000) وهو اقل من (0.05) وبالتالي فان قيمة (F) دالة إحصائياً، اي يمكننا القول ان معادلة الانحدار ككل دالة احصائياً عند مستوى دلالة اقل من (0.05).

(4) كما ان مخرجات البرنامج توفر قيمة معامل التحديد كما في الجدول (3-6)

جدول (3-6)

ملخص نتائج تحليل الانحدار

R (1)	R Square (R^2) (2)	Adjusted R ($\overline{R}^2$) Square (3)	Std. Error of the Estimate (4)
.801	.641	0.614	2.8345

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

ونلاحظ في جدول (3-6) ملخص لنتائج تحليل الانحدار يظهر العمود (1) معامل الارتباط المتعدد (R= 0.801) بين المتغير التابع والمتغير المستقل وهو نفسه معامل بيرسون لان النموذج يحوي على متغير مستقل واحد، فهو يدل على ان هناك علاقة عالية بين المتغير المستقل (خبرة العامل) والمتغير التابع (جودة المنتج)، ويشير العمود (2) الى معامل التحديد الذي يفسر نتائج النموذج، وقيمته تساوي(R^2 =0.641)وتعني ان النموذج المقدر يعبر عن (64)% من البيانات أي ان المتغيرالمستقل يفسر(64)% من تباين المتغير التابع، وفي العمود (3)تم حساب معامل التحديد المعدل للنموذج والذي يساوي $(\overline{R}^2 = 0.614)$ ويستخدم للغرض نفسه المذكوراً آنفاً ولكن بصورة ادق.

الخواص الحسابية لمقدرات المربعات الصغرى الاعتيادية:

١- يمر الخط المستقيم المقدر من نقطة متوسطات العينة لـ Y و X :

البرهان: $\because \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \quad \Rightarrow \quad \overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$

$(\overline{Y}, \overline{X})$ إحداثيات نقطة المتوسط.

٢- القيمة المتوسطة لقيم Y المقدرة $(\overline{\hat{Y}})$ تساوي القيمة المتوسطة لقيم Y الفعلية $(\overline{Y})$:

البرهان:

$$\overline{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} = (\overline{Y} - \hat{\beta}_1 \overline{X}) + \hat{\beta}_1 \overline{X}$$
$$= \overline{Y}$$

٣- مجموع البواقي يساوي صفراً:

البرهان:

$$\Sigma e_i = \Sigma (Y_i - \hat{Y}_i)$$

$$= \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

تمثل المعادلة الاولى من المعادلات الطبيعية.

$$\Rightarrow \quad \bar{e} = 0$$

٤- البواقي e_i لا ترتبط بـ X_i :

البرهان

$$\Sigma X_i e_i = \Sigma X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$= \Sigma X_i Y_i - \hat{\beta}_0 \Sigma X_i - \hat{\beta}_1 \Sigma X_i^{\,2}$$

تمثل المعادلة الثانية من المعادلات الطبيعية.

$$= 0$$

٥- البواقي لا ترتبط بالقيمة المقدرة لـ Y_i :

البرهان:

$$\mathrm{E}(e_i \hat{Y}_i) = \mathrm{E} e \hat{y}_i$$

أو :

$$\Sigma e_i \hat{y}_i = \Sigma e_i (\hat{\beta}_1 x_i) = \hat{\beta}_1 \Sigma x_i e_i = 0$$

العلاقة بين t و F $(F = t^2)$:

البرهان:

$$F = \frac{\hat{\beta}_1^2 \Sigma x_i^2}{\Sigma e_i^2 / (n-2)} = \frac{\hat{\beta}_1^2 \Sigma x_i^2}{\hat{\sigma}^2}$$

$$= \frac{\hat{\beta}_1^2}{\hat{\sigma}^2 / S_{XX}} = \frac{\hat{\beta}_1^2}{v(\hat{\beta}_1)}$$

$$= \left(\frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)} \right)^2 = t^2 \quad . . . \quad (21-3)$$

العلاقة بين F و R^2:

$$\left[F = \frac{R^2(n-2)}{1-R^2} \right]$$

البرهان:

$$F = \frac{ESS}{RSS/(n-2)} = \frac{ESS(n-2)}{RSS}$$

$$= \frac{\left(\frac{ESS}{TSS} \right)(n-2)}{\frac{RSS}{TSS}} = \frac{R^2(n-2)}{1-R^2} \quad . . . \quad (22-3)$$

العلاقة بينt و R^2:

$$t = \sqrt{F} = \sqrt{\frac{R^2(n-2)}{1-R^2}} \quad . . . \quad (23-3)$$

(8-3)الترجيح ووحدات القياس:Scaling and unites of Measurement

من أجل تبسيط الحسابات في بعض الحالات أو لتسهيل تفسير النتائج في حالات أخرى يتم استخدام وحدات قياس معقولة.

افترض ان وحدات القياس للمتغير X و Y في علاقة الانحدار $Y = \beta_0 + \beta_1 X + u$ مقاسة ببليون الوحدات. وباحث آخر استخدم بيانات عن X و Y مقاسة بمليون الوحدات، فهل ان نتائج الانحدار تكون متماثلة في الحالتين؟

ويمكن صياغة السؤال بالشكل الآخر التالي: هل ان قياس الوحدات للمتغيرين X و Y تولد اختلافات في نتائج الانحدار من ناحية تقدير المعلمات او الاستدلال حول معنوية معلماتها. وللإجابة عن السؤال افترض:

$$Y_i^* = W_1 Y_i \quad \& \quad X_i^* = W_2 X_i$$

حيث W_1 و W_2 هي أوزان التحويل (ثوابت) قد تكون متساوية او مختلفة.

مثلاً: X_i و Y_i مقاسة بالبليون من الوحدات، وأراد الباحث تحويلها الى (مليون من الوحدات)

وبذلك فان $Y_i^* = 1000Y_i \quad , \quad X_i^* = 1000X_i$ (W1 =W2 =1000)

بشكل عام:

الان افترض الانحدار باستخدام المتغيرات Y_i^*, X_i^* :

$$Y_i^* = \beta_0^* + \beta_1^* X_i^* + u_i^* \qquad . . . \qquad (24-3)$$

حيث ان: $Y_i^* = W_1 Y_i \quad , \quad X_i^* = W_2 X_i$

والآن نسعى لإيجاد العلاقات بين الآتي:

(1) $\hat{\beta}_0 \quad , \quad \hat{\beta}_0^*$.

(2) $\hat{\beta}_1 \quad , \quad \hat{\beta}_1^*$.

(3) $\text{var}(\hat{\beta}_0) \quad , \quad \text{var}(\hat{\beta}_0^*)$.

(4) $\text{var}(\hat{\beta}_1) \quad , \quad \text{var}(\hat{\beta}_1^*)$.

(5) $\hat{\sigma}^2 \quad , \quad \hat{\sigma}^{*2}$

(6) $r_{XY}^{\ 2} \quad , \quad r_{X^*Y^*}^{\ 2}$.

باستخدام قوانين المربعات الصغرى لتقدير المعلمات:

• المعلمة β_0 : $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \Rightarrow \quad \hat{\beta}_0^* = \bar{Y}^* - \hat{\beta}_1^* \bar{X}^*$

$$= W_1 \bar{Y} - \frac{W_1}{W_2} \hat{\beta}_1 \cdot (W_2 \bar{X}) = W_1(\bar{Y} - \hat{\beta}_1 \bar{X}) = W_1 \hat{\beta}_0$$

$$\hat{\beta}_0^* = W_1 \hat{\beta}_0 \qquad . . . \qquad (25-3)$$

$$\hat{\beta}_1^* = \frac{\Sigma x^* y^*}{\Sigma x^{*2}} = \frac{\Sigma(w_2 x)(w_1 y)}{\Sigma(w_2 x)^2} = \frac{\Sigma w_2 w_1 (xy)}{\Sigma w_2^2 x^2} = \frac{w_1}{w_2} \frac{\Sigma xy}{\Sigma x^2}$$

$$\hat{\beta}_1^* = \frac{w_1}{w_2} \cdot \hat{\beta}_1 \qquad . . . \qquad (26-3)$$

وهكذا يمكن البرهنة على: $\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2$

$$\text{var}(\hat{\beta}_0^*) = w_1^2 \, \text{var}(\hat{\beta}_0) \qquad . . . \qquad (27-3)$$

$$\text{var}(\hat{\beta}_1^*) = \left(\frac{w_1}{w_2}\right)^2 \text{var}(\hat{\beta}_1) \qquad ... \qquad (28-3)$$

$$r_{XY}^2 = r_{X^*Y^*}^2 \qquad ... \qquad (29-3)$$

خلاصة القول: بمعرفة الأوزان (w_i) يمكن عملياً اختيار وحدات القياس المناسبة سواء للحسابات أو للتفسير.

(1) فاذا كانت وحدات قياس المتغير X والمتغير Y متساوية ($w_1=w_2=w$) فان معلمة الانحدار β_1 وانحرافه المعياري s.e (β_1) لا تتغير في كلا معادلتي الانحدار باستخدام المتغيرات (Y , X) أو (Y^* , X^*). في حين يكون المقطع الصادي وانحرافه المعياري مرجحاً بالوزن w .

(2) اذا كانت وحدات X لم تتغير أي ($w_2=1$) في حين وحدات قياس Y تتغير بالوزن (w_1) فان الميل (β_1) والمقطع الصادي (β_0) وانحرافهما المعياري يتم ترجيحها بالوزن (w_1) .

(3) اذا وحدات Y لم تتغير أي ($w_1=1$) في حين وحدات قياس X تتغير(w_2) فان الميل (β_1) وانحرافه المعياري s.e (β_1) يتم ترجيحها بالوزن $(\frac{1}{w_2})$ في حين يبقى المقطع الصادي (β_0) وانحرافه المعياري s.e (β_0) لايتأثر.

(4) ان تحويل X و Y الى X^* و Y^* لن يؤثر في صفات المقدرات بطريقة OLS .

مثال(3-18) : افترض باحث استخدام بيانات الجدول (2-1) وأراد أن يحسب (Y) بالكيلوغرامات وتبقى X بالهكتار. فما هي معادلة التقدير.

الجواب: $w_2=1$ بينما $=1000w_1$ ، $Y_i^* = 1000 y_i \quad X_i^* = X_i$

وحيث ان معادلة الانحدار باستخدام X و Y هي:

$$\hat{Y}_i = 35.35 + 2.564 X_i$$

$$\hat{Y}_i^* = 35350 + 2564 X_i^*$$

$$s.e: \ (24668.4) \quad (260.768)$$

وكذلك الانحراف المعياري

$$\hat{\sigma}^{*2} = 2308.3802$$

ويبقى تفسير المعلمات.

$(\hat{\beta}_1)$: زيادة المساحة المزروعة بمقدار هكتار واحد تؤدي الى زيادة الانتاج بمقدار (2564) كغم أو (2.564) ألف كغم.

أسـئـلـة الـفـصـل الـثـالـث:

س١: اجب باختصار عما يأتي:

أ. كيف نحدد فرضية العدم

ب. الخطأ من النوع الاول (Type I error)

ج. الخطأ من النوع الثاني(Type II error)

د. رفض فرضية العدم: $H_0 : \beta_1 = 0$

هـ. قبول فرضية العدم: $H_0 : \beta_0 = 0$

و. حدود الثقة للمعلمات المقدرة.

ز. الربط الجوهري بين مجال الثقة واختبار المعنوية.

ح. الفرق بين متوسط الاستجابة $E(Y_t / X = X_f)$ وبين التنبؤ بالاستجابة الجديدة (Y_{fnew})

ط. التفسير البديهي لمفهوم درجات الحرية.

ي. ان R2 قد يكون مقاساً غير جيد للدلالة على ملاءمة النموذج للبيانات.

ك. تأثير الاختلاف في وحدات قياس المتغيرين X و Y على نتائج الانحدار.

س2: صحح الخطأ ان وجد في العبارات التالية:

١. ان قبول فرضية العدم او رفضها يتم على اساس ان قيمة المعلمة الحقيقية تختلف عن الصفر.

٢. ان الخطأ من النوع الاول يكافئ الخطأ من النوع الثاني.

٣. اذا اردنا اثبات ان الميل لخط الانحدار(β_1) لا يساوي صفراً فتشكل فرضية العدم: $H_0 : \beta_1 = 0$ ، اما اذا كنا نسعى الى اثبات ان ميل الانحدار للمجتمع موجباً فتكون فرضية العدم $H_0 : \beta_1 \geq 0$.

٤. في حالة الاختبار ذي الطرفين فان معنوية المعلمة المقدرة باستخدام مستوى دلالة 5% يحدد اذا كانت قيمة المعلمة المقدرة اكبر من $\left(\frac{1}{2}\right)$ حجم الخطأ المعياري المقدر.

٥. قبول فرضية العدم $H_0 : \beta_1 > 0$ اذا كانت القيمة الحسابية ضمن منطقة القبول

$$t^*_{\hat{\beta}_1} \geq (n-2, \frac{\alpha}{2})$$

٦. اذا كانت القيمة المختبر حولها للمعلمة تقع ضمن حدود الثقة فذلك دليل على رفض فرضية العدم والعكس صحيح.

٧. تباين متوسط الاستجابة تعادل تباين خطأ التنبؤ.

٨. ان فترة الثقة للقيمة التنبؤية الجديدة تزداد مع ابتعاد القيمة التنبؤية لـ X عن نقطة المركز.

٩. الاحصاءة F هي النسبة لمتغيرين مستقلين ذو توزيع χ^2 .

١٠. ان استخدام جدول تحليل التباين يصلح لاختبار معنوية المعلمات المقدرة بشكل عام.

١١. معامل الارتباط البسيط هو مقياس لدراسة الترابط بين متغيرين ويعكس العلاقة السببية بينهما.

١٢. ان دقة المعالم المقدرة تعتمد على تحقيق الفروض والصيغة المستخدمة لاختبار هذه الفروض هو رسم الانتشار.

١٣. اذا كانت فترة الثقة للمعلمة β هي: $(-0.3 \le \beta \le 0.9)$ فان المعلمة β تكون معنوية.

س3: اشتق التوزيع المناسب لمتوسط الاستجابة عند مستوى معلوم للمتغير $X(X_0)$. وقارن بينه وبين التوزيع الاحتمالي لـ Y_f التنبؤية الجديدة ؟

س4: بين أهم ملامح الاختلاف بين تحليل الارتباط وتحليل الانحدار؟

س5: برهن ان معامل التحديد يعكس مربع معامل الارتباط البسيط بين المتغيرين X و Y ؟

س6: اكتب جدول تحليل التباين (ANOVA) معتمداً على معامل التحديد وحجم العين ؟

س7: معادلة الانحدار المقدرة:

$$\hat{Y}_i = -200 + 2.3X_i \qquad i = 1,\cdots,20$$
$$t \;:\; \qquad (2.9)$$

احسب معامل التحديد وفسر دلالته ؟

س٨: اذا علمت ان معادلة التقدير:

$$\hat{Y}_i = 30.5 - 0.8X_i \qquad i = 1,\cdots,18$$
$$s.e \;:\; \qquad (0.04)$$

$$\Sigma X = 62 \quad , \quad R^2 = 0.42 \quad , \quad \Sigma y^2 = 490$$

تنبأ عن قيمة Y عندما X = 20 ثم اختبر معنويتها ؟

س٩: حدد حجم العينة المستخدمة اذا علمت:

$$\hat{Y}_i = 12.5 + 2.3X_i$$
$$t \;:\; \qquad (0.3) \qquad\qquad R^2 = 0.72$$

س١٠:(1)اختبر معنوية العلاقة الخطية بين Y و X اذا علمت: $Y' = [8 \quad 9 \quad 12 \quad 17 \quad 21]$ وان $R^2 = 0.75$.

(٢) احسب تباين خطأ التقدير.

س11: الجدول التالي يمثل عدد ساعات المطالعة العلمية اليومي (X) والمعدل العام (Y) لعينة من الطلبة:

X	2	5	6	4
Y	60	80	75	65

١. قدر معدل طالب يدرس (3) ساعات . واحسب مجال ثقة التقدير ؟

٢. احسب قيمة معامل التحديد وفسر دلالته الإحصائية. ثم كون جدول تحليل التباين بالاعتماد على معامل التحديد ؟

س12: اختبر الفرضية $1H_0 : 2\beta_1 >$ إذا علمت:

$$\hat{Y}_i = 22.01 + 0.973X_i \qquad i = 1, \cdots, 20$$

$$\Sigma x^2 = 55 \quad , \quad \Sigma y^2 = 64$$

س13: باحث يروم استخدام الصيغة الخطية لتقدير العلاقة بين Y و X وقد التبس عليه أي النموذجين يستخدم $Y_i = \beta_1 X_i + u_i$ أم $Y = \beta_0 + \beta_1 X_i + u_i$

ما مقترحاتك لإزالة هذا اللبس ؟

س14: اعتماداً على الخلاصة الإحصائية التالية: $\hat{Y}_i = 2.6 + 1.5X_i \quad , \quad i = 1, \cdots, 20$

$$\bar{X} = 5 \quad , \quad \Sigma Y = 120 \quad , \quad \Sigma Y^2 = 903.5 \quad , \quad \Sigma X^2 = 568.9$$

اختبر معنوية متوسط الاستجابة عندما X = 10 باستخدام مستوى دلالة 5% ؟

الـفـصـل الـرابـع

الانـحـدار الـخـطي الـمـتـعـدد Multiple Linear Regression

يعد نموذج الانحدار المتعدد (النموذج الخطي العام General Linear Model) الامتداد الطبيعي والمنطقي للنموذج الخطي بمتغيرين. ففي حالة استخدام k من المتغيرات المستقلة $(X_1,X_2,...,X_k)$ لتفسير تباين المتغير المعتمد (التابع) Y في معادلة الانحدار، فان جميع المفاهيم في هذه الحالة تتشابه مع حالة نموذج الانحدار البسيط. غير ان تعدد المتغيرات المستقلة تجعل التعامل مع طرائق الجبر الخطي (جبر المصفوفات) هي المستخدمة لتقدير واختبار وتحليل نماذج الانحدار المتعدد. وبذلك يمكن تعميمها وتطبيقها على حالات المتغيرين، او ثلاثة المتغيرات او أي عدد من المتغيرات بشرط لا يفوق عدد المتغيرات على عدد المشاهدات المستخدمة للتقدير.

(1-4)النموذج الخطي العام (من خلال نقطة الأصل) Through the origin

نفترض ان المتغير المعتمد Y دالة خطية بدلالة (k) من المتغيرات المستقلة:

$(X_1,X_2,...,X_k)$ وبذلك يصاغ نموذج الانحدار الخطي العام على وفق العلاقة:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i \qquad ... \qquad (1-4)$$

$$; i = 1,2,...,n$$

حيث ان:

$\beta_j \ , j = 1,...,k$: تمثل معالم النموذج الجزئية التي تقيس استجابة المتغير التابع للمتغير المستقل المعني مع جعل بقية المتغيرات المستقلة ثابتة. وهي تمثل مقدار التغير في Y للتغير بوحدة واحدة من X_i مع ثبات بقية المتغيرات المستقلة.

$u_i \ , i = 1,...,n$: تمثل قيم المتغير العشوائي المجهولة.

ولعينة من (n) من المشاهدات، فان العلاقة (1-4) تتحقق لكل مشاهدة فيتكون n من المعادلات وعلى وفق الآتي:

$$Y_i = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + ... + \beta_k X_{k1} + u_1$$

$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + ... + \beta_k X_{k2} + u_2$$

.

.

.

$$Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + ... + \beta_k X_{kn} + u_n$$

وباستخدام المصفوفات يحول نظام المعادلات على وفق:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & ... & X_{k1} \\ 1 & X_{12} & X_{22} & ... & X_{k2} \\ . \\ . \\ . \\ 1 & X_{1n} & X_{2n} & ... & X_{kn} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{pmatrix}$$

وباختصار:

$$Y = X\beta + u \quad ... \quad (2-4)$$

حيث ان:

Y: متجه عمودي لمشاهدات المتغير المعتمد وبترتيب $(n \times 1)$.

X: مصفوفة تحتوي مشاهدات المتغيرات المستقلة X_1 ، X_2 ، . . . . ، X_k وعمودها الاول متغير وهمي للدلالة على المقطع الصادي. وترتيب المصفوفة $n \times (k+1)$. حيث ان الصف الأول من هذه المصفوفة هو: $(1 \quad X_{11} \quad X_{21} ... X_{k1})$. وتسمى المصفوفة بمصفوفة المعلومات (data matrix).

β : متجه عمودي يحتوي معالم النموذج الخطي المراد تقديرها. وبترتيب $(k+1) \times 1$.

u : متجه عمودي يحوي قيم المتغير العشوائي المجهولة. وبترتيب $(n \times 1)$.

الطرف الأيمن في العلاقة (4-2) يمثل الجزء المحدد بالمتغيرات المستقلة (Xβ)مضاف اليها الجزء الاحتمالي العشوائي (u) .

مثال(4-1): ولتوضيح التمثيل بالصيغة المصفوفية في حالة متغيرين Y و X ولسبعة مشاهدات كما في الجدول (4-1) .

جدول (4-1)

Y	70	65	90	95	١١٠	115	120
X	80	100	120	140	160	180	200

فان معادلة الانحدار بصيغة المصفوفات هي:

$$\begin{bmatrix} 70 \\ 65 \\ 90 \\ 90 \\ 110 \\ 115 \\ 120 \end{bmatrix} = \begin{bmatrix} 1 & 80 \\ 1 & 100 \\ 1 & 120 \\ 1 & 140 \\ 1 & 160 \\ 1 & 180 \\ 1 & 200 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix}$$

$$\underset{(7\times1)}{Y} = \underset{(7\times2)}{X} \ \underset{(2\times1)}{\beta} + \underset{(7\times1)}{u}$$

كما في حالة الانحدار الخطي البسيط. يكون الهدف الرئيس بعد جمع البيانات وتمثيل الصيغة الدالية يكون الهدف اللاحق هو تقدير معلمات النموذج المتمثل بالعلاقة (4-1) على وفق الصيغة المصفوفية العامة (4-2) ، يكون المطلوب إيجاد تقدير المتجه β. ويتم ذلك على وفق طريقة (OLS) أو طريقة الإمكان الأعظم .ML . وحيث ان الطريقتين تعطي نتائج متماثلة فيما يخص المعلمات المقدرة β لذا سيتم التركيز على طريقة المربعات الصغرى OLS .

(4-2)فرضيات نموذج الانحدار الخطي بصيغة المصفوفات Assumptions of classical linear regression model in matrix form

•الفرضية الأولى: متوسط متجه المتغير العشوائي لكل عنصر من عناصره، يساوي صفراً. أي:

$$\mathrm{E}(u_i) = 0 \quad , \ i = 1,\dots,n$$

وبالصيغة المصفوفية:

$$\mathrm{E}(u) = \mathrm{E}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ . \\ . \\ . \\ u_n \end{pmatrix} = \begin{bmatrix} \mathrm{E}(u_1) \\ \mathrm{E}(u_2) \\ \mathrm{E}(u_3) \\ . \\ . \\ . \\ \mathrm{E}(u_n) \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \quad ... \quad (3-4)$$

•الفرضية الثانية: تباين المتغير العشوائي متجانس، قيمة التباين متساوية لكل مشاهدة. أي:

$$\mathrm{E}(u_i u_j) = \mathrm{var}(u_i) = \sigma^2 \quad , \quad i = j$$ (فرضية تجانس التباينHomoscedasticity)

•الفرضية الثالثة: التباين المشترك لأي مشاهدتين من مشاهدات المتغير العشوائي، تساوي صفراً أي:

$$E(u_i u_j) = 0 \quad , \quad i \neq j$$

وتسمى فرضية عدم وجود ارتباط ذاتي (No serial correlation)

وبالصيغة المصفوفية يمكن دمج الفرضية الثانية والثالثة كآلاتي:

$$E(uu') = E\begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{pmatrix} \begin{pmatrix} u_1 & u_2 & . & . & . & u_n \end{pmatrix}$$

$$E(uu') = E\begin{pmatrix} u_1^2 & u_1 u_2 & u_1 u_3 & . & . & . & u_1 u_n \\ u_2 u_1 & u_2^2 & u_2 u_3 & . & . & . & u_2 u_n \\ . & . & . & & & & . \\ . & . & . & & & & . \\ . & . & . & & & & . \\ u_n u_1 & u_n u_2 & u_n u_3 & . & . & & u_n^2 \end{pmatrix}$$

$$
= \begin{pmatrix}
Eu_1^2 & Eu_1u_2 & Eu_1u_3 & . & . & . & Eu_1u_n \\
Eu_2u_1 & Eu_2^2 & Eu_2u_3 & . & . & . & Eu_2u_n \\
. & . & . & & . & & \\
. & . & . & & . & & \\
. & . & . & & . & & \\
Eu_nu_1 & Eu_nu_2 & Eu_nu_3 & . & . & & Eu_n^2
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sigma^2 & 0 & \cdots & 0 \\
0 & \sigma^2 & \cdots & 0 \\
\vdots & \cdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma^2
\end{pmatrix} = \sigma^2 \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \cdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix}
$$

$$
E(uu') = \sigma^2 I_n \quad ... \quad (4-4)
$$

والمصفوفة الممثلة بالعلاقة (4-4) تمثل مصفوفة التباين. والتباين المشترك للمتغير العشوائي u بالصيغة المصفوفية، عناصر القطر الرئيس (main diagonal) تمثل تباين مشاهدات المتغير العشوائي. في حين عناصر المثلث العلوي تساوي عناصر المثلث السفلي لأنها تمثل التباين المشترك لمشاهدات المتغير العشوائي. وغني عن الذكر المصفوفة تكون متماثلة (symmetric).

الفرضية الرابعة: المتغيرات $X_1, X_2, ..., X_k$ متغيرات ثابتة للعينة المختارة فهي غير عشوائية بمعنى ان المتغيرات التوضيحية تكون متغيرات خارجية (Exogenous) ، وبذلك فان المصفوفة X ذات الترتيب $n \times (k+1)$ غير عشوائية. بمعنى تحتوي على أرقام ثابتة.

الفرضية الخامسة: لا توجد علاقة خطية تامة بين المتغيرات التي تمثل المصفوفة X . بمعنى لا وجود للتعدد الخطي. (Multicollinearity).

ويمكن تجزئة الفرضية كالآتي:

$$r_{x_0 x_j} = 0 \quad \forall \quad j = 1,2,\dots,k \quad (١)$$

وتقرأ باستقلال العمود الأول (والذي مشاهداته واحد) مع أي من أعمدة المصفوفة الأخرى.
وعملياً يعنـي وجـود تغييـرات واضـحة وملموسـة فـي مشـاهدات كـل متغيـر مـن المتغيـرات المسـتقلة (التوضيحية) X_1، . . . ، X_k .

$$r_{x_i x_j} = 0 \quad \forall\, i \neq j \quad , \quad i, j = 1,2,\dots,k \quad (٢)$$

وبذلك يفترض استقلال المتغيرات التوضيحية عن بعضها الآخر. فالأعمدة الثاني والثالث ...... والعمود (k+1) مستقلة عن بعضها الآخر.
ويمكن عرض هذه الفرضية بالصيغة المصفوفية:

$$\rho(X) = k+1 \;\langle\; n \qquad . . . \qquad (5-4)$$

رتبة مصفوفة المعلومات (X) تامة من ناحية الأعمدة وهي أقل من عدد المشاهدات (n) .

الفرضية السادسة: يفترض ان مشاهدات المتغير العشوائي u تتوزع بشكل متماثل أي لها توزيع طبيعي (Normal) . وهذه الفرضية مهمة لمرحلة اختبار الفرضيات في الفصل الخامس.
وبصيغة المصفوفات:
ان متجه المتغير العشوائي u له توزيع طبيعي متعدد. u ~ N
وبدمج الفرضيات (2) و (3) و (6) يمكن صياغتها مصفوفياً كالآتي:
$u \sim N(0, \sigma^2 I_n)$ . . . (6-4)

(3-4)<u>تقدير المعلمات بطريقة المربعات الصغرى الاعتيادية</u>:" <u>Parameter Estimation by OLS</u> "

ان المعيار الذي تستند إليه طريقة المربعات الصغرى الاعتيادية هو جعل مجموع مربعات الخطأ (البواقي) أقل ما يمكن. وحيث ان البواقي تمثل قيم y الحقيقية مطروحاً منها القيم التقديرية $(\hat{y})$. أي:

$$e_i = Y_i - \hat{Y}_i \quad \forall \quad i = 1,2,\dots,n$$

وباستخدام المصفوفات:

$$\begin{aligned} e &= Y - \hat{Y} \\ &= Y - X\hat{\beta} \end{aligned}$$

وبذلك فان دالة الهدف بالصيغة المصفوفية:

$$\sum_{i=1}^{n} e_i^2 = e'e$$

$$= (e_1 \ e_2 \ . \ . \ . \ e_n) \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + \ . \ . \ . \ + e_n^2$$

$$e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= Y'Y - 2\hat{\beta}'X'Y - \hat{\beta}'X'X\hat{\beta} \qquad . \ . \ . \qquad (7-4)$$

حيث ان $(\hat{\beta}'X'Y)$ متجه بترتيب (scalar)(1×1) أو عدد حقيقي وبذلك

$$Y'X\hat{\beta} = (\hat{\beta}'X'Y)' = \hat{\beta}'X'Y$$

العلاقة (4-7) تمثل دالة الهدف التي نسعى لتصغيرها وهي دالة بدلالة متجه المعلمات $\hat{\beta}$.

وبتطبيق الشرط الضروري:

$$\frac{\partial e'e}{\partial \hat{\beta}} = 0$$

تعطي (k+1) من المعادلات والتي تمثل المعادلات الطبيعية.

$$\frac{\partial e'e}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \qquad . \ . \ . \qquad (8-4)$$

وبحل نظام المعادلات (4-6) بالصيغة المصفوفية:

$$X'X\hat{\beta} = X'Y \qquad . \ . \ . \qquad (9-4)$$

نظام المعادلات (4-9) يمثل المعادلات الطبيعية كالآتي:

$$n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_{1i} + \hat{\beta}_2 \Sigma X_{2i} + . . . + \hat{\beta}_k \Sigma X_{ki} = \Sigma Y_i$$

$$\hat{\beta}_0 \Sigma X_{1i} + \hat{\beta}_1 \Sigma X_{1i}^2 + \hat{\beta}_2 \Sigma X_{1i} X_{2i} + . . . + \hat{\beta}_k \Sigma X_{1i} X_{ki} = \Sigma X_{1i} Y_i$$

$$\vdots$$

$$\hat{\beta}_0 \Sigma X_{ki} + \hat{\beta}_1 \Sigma X_{ki} X_{1i} + \hat{\beta}_2 \Sigma X_{ki} X_{2i} + . . . + \hat{\beta}_k \Sigma X_{ki}^2 = \Sigma X_{ki} Y_i$$

وبالصيغة المصفوفية:

$$\begin{bmatrix} n & \Sigma X_{1i} & \Sigma X_{2i} & \cdots & \Sigma X_{ki} \\ \Sigma X_{1i} & \Sigma X_{1i}^2 & \Sigma X_{1i} X_{2i} & \cdots & \Sigma X_{1i} X_{ki} \\ \vdots \\ \Sigma X_{ki} & \Sigma X_{ki} X_{1i} & \Sigma X_{ki} X_{2i} & \cdots & \Sigma X_{ki}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \vdots \\ \Sigma X_k Y \end{bmatrix}$$

$$\underset{(k+1)\times(k+1)}{X'X} \; \underset{(k+1)\times 1}{\hat{\beta}} = \underset{(k+1)\times 1}{X'Y}$$

حيث ان:

المصفوفة $X'X$ تسمى مصفوفة فيشر للمعلومات، هي بترتيب $(k+1)\times(k+1)$ وهي مصفوفة مربعة (square) ومتماثلة (symmetric) وغير شاذة (nonsingular) لانها ذات رتبة تامة من ناحية الأعمدة على وفق الفرضية الخامسة { العلاقة (4-5)} وبذلك فان معكوسها موجود $(X'X)^{-1}$. وبضرب طرفي العلاقة (4-9) بـ $(X'X)^{-1}$ وتبسيط الطرفين يتم الحصول على قانون المربعات الصغرى بالصيغة المصفوفية:

$$\underset{(k+1)\times 1}{\hat{\beta}} = \underset{(k+1)\times(k+1)}{(X'X)^{-1}} \; \underset{(k+1)\times 1}{X'Y} \qquad \ldots \qquad (10-4)$$

ونرمز للمصفوفة $(X'X)^{-1}$ بالرمز c معكوس مصفقوفة فيشر.

ولتوضيح استخدام المصفوفات، يتم استخدام معلومات المثال (4-1) كحالة خاصة للانحدار الخطي البسيط.

$$\Sigma Y = 665 \;,\; \Sigma X = 980 \;,\; \Sigma X^2 = 148400 \;,\; \Sigma XY = 98500$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$(X'X) = \begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix} = \begin{pmatrix} 7 & 980 \\ 980 & 148400 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 665 \\ 98500 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 1.89 & -0.013 \\ -0.013 & 0.000089 \end{pmatrix} = C = \begin{pmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{pmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} -23.65 \\ 0.12 \end{bmatrix}$$

مثال (2-4) : حالة ثلاث متغيرات:

بيانات افتراضية خمس مشاهدات لقيم Y و X_1 و X_2 .

جدول (2-4)

Y	3	1	8	3	5	∑20
X_1	3	1	5	2	4	∑15
X_2	5	4	6	4	6	∑25

$$\Sigma X_1^2 = 55 \ , \ \Sigma X_2^2 = 129 \ , \ \Sigma X_1 Y = 76 \ , \ \Sigma X_2 Y = 109$$

تمثيل النموذج بصيغة المصفوفات

$$\begin{pmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{pmatrix} = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix}$$

المعادلات الطبيعية:

$$X'X \ \hat{\beta} = X'Y$$

$$\begin{bmatrix} 5 & 15 & 25 \\ 15 & 55 & 81 \\ 25 & 81 & 129 \end{bmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 20 \\ 76 \\ 109 \end{pmatrix}$$

ولإيجاد المعلمات المقدرة يتطلب:

١- حساب $(X'X)^{-1}$.

٢- أو باستخدام طريقة كريمر باستخدام المحددات.

٣- أو باستخدام تحويل المصفوفة $(X'X)$ الى المصفوفة مثلثية.

باستخدام طريقة (Elimination).

وسيتم استخدام الطريقة (2) (طريقة كريمر) :

$$\hat{\beta}_0 = \frac{\begin{vmatrix} 20 & 15 & 25 \\ 76 & 55 & 81 \\ 109 & 81 & 129 \end{vmatrix}}{\begin{vmatrix} 5 & 15 & 25 \\ 15 & 55 & 81 \\ 25 & 81 & 129 \end{vmatrix}} = 4$$

$$\hat{\beta}_1 = \frac{\begin{vmatrix} 5 & 20 & 25 \\ 15 & 76 & 81 \\ 25 & 109 & 129 \end{vmatrix}}{\begin{vmatrix} 5 & 15 & 25 \\ 15 & 55 & 81 \\ 25 & 81 & 129 \end{vmatrix}} = 2.5 \quad , \quad \hat{\beta}_2 = \frac{\begin{vmatrix} 5 & 15 & 20 \\ 15 & 55 & 76 \\ 25 & 81 & 109 \end{vmatrix}}{\begin{vmatrix} 5 & 15 & 25 \\ 15 & 55 & 81 \\ 25 & 81 & 129 \end{vmatrix}} = -1.5$$

اذن معادلة الانحدار التقديرية:

$$\hat{Y} = 4 + 2.5X_1 - 1.5X_2$$

أو:

$$Y = 4 + 2.5X_1 - 1.5X_2 + e$$

(4-4)معنى معاملات الانحدار في حالة ثلاثة متغيرات: $E(Y_i / X_{1i}, X_{2i})$:

$$E(Y_i / X_{1i}, X_{2i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

β_1 : وهي معلمة تقيس التغير في متوسط قيم Y . لوحدة تغير في X_1 ، مع إبقاء X_2 ثابتاً . وهي تقيس ميل Y نسبة إلى X_1 مع إبقاء X_2 ثابتاً.

ورياضياً فان المشتقة الجزئية لـ Y نسبة إلى X_1 $(\frac{\partial Y}{\partial X_1})$. وهي تقيس الأثر المباشر (direct) أو الصافي (net) لوحدة التغير في X_1 على متوسط قيم Y صافي لـ X_2 . وكذلك الحال بالنسبة لـ β_2 فهي تقيس الأثر المباشر فقط X_2 على Y. ويمكن قياس الأثر المباشر بالطريقة التي تتركز بإزالة أثر X_2 على وفق الخطوات التالية:

١- نجري انحدار Y على X_2 فقط: $Y = b_0 + b_{12} X_2 + e_1$

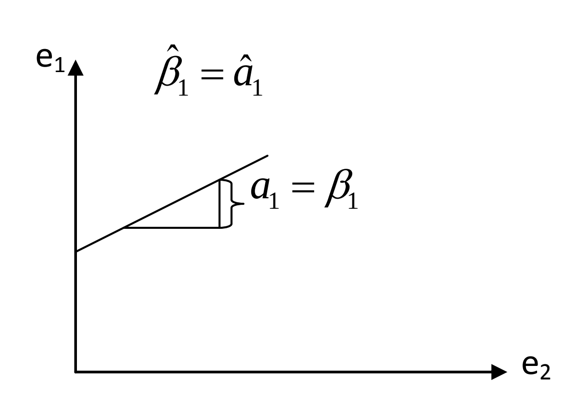
٢- انحدار X_1 على X_2 فقط: $X_1 = b_1 + b_{12} X_2 + e_2$

حيث ان e_1 : تمثل قيم Y بعد إزالة الأثر الخطي لـ X_2 على Y.

e_2 : تمثل قيم X_1بعد إزالة الأثر الخطي لـ X_2 على X_1.

٣- نجري انحدار e_1 علىe_2 : $e_1 = a_0 + a_1 e_2 + e_3$

وبذلك يكون $(\hat{a}_1)$ هو التقدير الحقيقي ، اذ يقيس الأثر الصافي لوحدة التغير في X_1 على Y . او الميل الحقيقي لـ Y بالنسبة لـ X_1 وهي بذلك :



(5-4)صفات متجه المعلمات المقدرة Properties of estimated parameter vector

1− المعلمات المقدرة تقديرات غير متحيزة لقيم المعلمات الحقيقية في المجتمع. لبرهنة هذه الصفة بالصيغة المصفوفية:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta+u) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u \quad ... \quad (11-4)\end{aligned}$$

وبأخذ التوقع للطرفين:

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(u)$$

$$E(\hat{\beta}) = \beta$$

حيث ان E(u) = 0 (الفرض الأول)

$$E\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \vdots\\ \hat{\beta}_k\end{pmatrix} = \begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{pmatrix}$$

2- المتجه $\hat{\beta}$ يعد تركيب خطي بدلالة المتجه Y

وحيث ان X ثابتة للعينة المختارة

اذن $(X'X)^{-1}X$ ثابتة أيضاً

نفرض $\Rightarrow (X'X)^{-1}X = A$

$$\therefore \hat{\beta} = AY$$

3− المعلمات المقدرة لها أقل تباين بموجب نظرية (Gauss Markov) .

وبذلك فان المقدرات بطريقة OLS هي أفضل مقدرات خطية غير منحازة او ما يطلق عليها اختصاراً (BLUE).

(6-4)مصفوفة التباين – والتباين المشترك لـ $\hat{\beta}$ Variance−Covariance Matrix.

ان مصفوفة التباين والتباين المشترك للمعلمات المقدرة تعد مهمة لإغراض الاستدلال الإحصائي حول المعلمات.

$$\text{var}-\text{cov}(\hat{\beta}) = E\left\{\left(\hat{\beta}-E(\beta)\right)\left(\hat{\beta}-E(\beta)\right)'\right\} \quad ... \quad (12-4)$$

وحيث ان:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta+u) \\ &= (X'X)^{-1}X'X\beta+(X'X)^{-1}X'u\end{aligned}$$

$$\hat{\beta} = \beta+(X'X)^{-1}X'u$$

$$\therefore \hat{\beta}-\beta = (X'X)^{-1}X'u$$

$$\begin{aligned}\text{var}-\text{cov}(\hat{\beta}) &= \text{E}\left[(X'X)^{-1}X'u\right]\left[u'X(X'X)^{-1}\right] \\ &= \text{E}\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right] \\ &= (X'X)^{-1}X'\left[\text{E}(uu')\right]\ X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}\end{aligned}$$

وباعتماد العلاقة (4-4):

$$\text{var}-\text{cov}(\hat{\beta}) = \sigma^2 C \qquad \ldots \qquad (13\text{-}4)$$

العلاقة (4-12) تبين عناصر التباين لمعلمات الانحدار فضلا عن التباين المشترك فيما بين المعلمات المقدرة والتي يمكن كتابتها بشكل صريح كالآتي:

$$\text{var}-\text{cov}(\hat{\beta}) = \text{var}-\text{cov}\begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k\end{pmatrix}$$

$$= \text{E}\begin{pmatrix}\hat{\beta}_0-\beta_0 \\ \hat{\beta}_1-\beta_1 \\ \vdots \\ \hat{\beta}_k-\beta_k\end{pmatrix}\left[(\hat{\beta}_0-\beta_0) \quad (\hat{\beta}_1-\beta_1) \quad . \; . \; . \quad (\hat{\beta}_k-\beta_k)\right]$$

$$\text{var}-\text{cov}(\hat{\beta}) = \begin{pmatrix} E(\hat{\beta}_0-\beta_0)^2 & E(\hat{\beta}_0-\beta_0)(\hat{\beta}_1-\beta_1) & . & . & . & E(\hat{\beta}_0-\beta_0)(\hat{\beta}_k-\beta_k) \\ & E(\hat{\beta}_1-\beta_1)^2 & . & . & . & E(\hat{\beta}_1-\beta_1)(\hat{\beta}_k-\beta_k) \\ & & . & & & \\ & & & . & & \\ & & & & . & \\ & & & & & E(\hat{\beta}_k-\beta_k)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0,\hat{\beta}_1) & . & . & . & \text{cov}(\hat{\beta}_0,\hat{\beta}_k) \\ \text{cov}(\hat{\beta}_1-\hat{\beta}_0) & \text{var}(\hat{\beta}_1) & . & . & . & \text{cov}(\hat{\beta}_1,\hat{\beta}_k) \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ \text{cov}(\hat{\beta}_k,\hat{\beta}_0) & \text{cov}(\hat{\beta}_k,\hat{\beta}_1) & . & . & . & \text{var}(\hat{\beta}_k) \end{pmatrix}$$

وبموجب العلاقة (4-13)

$$\text{var}-\text{cov}(\hat{\beta}) = \sigma^2 C$$

فان:

$$\text{var}(\hat{\beta}_i) = \sigma^2 c_{ii} \quad \forall i$$

$$\text{cov}(\hat{\beta}_i,\hat{\beta}_j) = \sigma^2 c_{ij} \quad \forall i \neq j \qquad \begin{matrix} i = 0,1,\dots,k \\ j = 0,1,\dots,k \end{matrix}$$

حيث ان c_{ii} تمثل عناصر القطر الرئيس للمصفوفة C.

c_{ij} تمثل عناصر المثلث العلوي (أو السفلي) للمصفوفة C.

$$\text{var}(\hat{\beta}_0) = \sigma^2 c_{00}$$

$$\text{var}(\hat{\beta}_1) = \sigma^2 c_{11}$$

$$\vdots$$

$$\text{var}(\hat{\beta}_k) = \sigma^2 c_{kk}$$

$$\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) = \sigma^2 c_{01}$$

$$\vdots$$

$$\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_k) = \sigma^2 c_{1k}$$

وهكذا.

وهكذا فان المعلمات المقدرة والتي تتصف بأنها غير متحيزة وإنها تركيب خطي بدلالة Y، وبذلك فانها تحمل توزيع Y نفسه، وحيث ان Y لها توزيع u نفسه الذي تم افتراضه بانه يتوزع توزيعاً طبيعياً فان المعلمات المقدرة تتوزع توزيعاً طبيعياً بمتوسط $\mathrm{E}(\hat{\beta}_i) = \beta_i$

$$\& \quad \mathrm{var-cov}(\hat{\beta}_i) = \sigma^2 c$$

والتي يمكن تلخيصها على وفق العلاقة التالية:

$$\hat{\beta} \sim \mathrm{N}(\beta, \sigma^2 C) \qquad \ldots \qquad (14\text{-}4)$$

$$\hat{\beta}_0 \sim \mathrm{N}(\beta_0, \sigma^2 c_{00})$$

$$\hat{\beta}_1 \sim \mathrm{N}(\beta_1, \sigma^2 c_{11})$$

$$\vdots$$

$$\hat{\beta}_k \sim \mathrm{N}(\beta_k, \sigma^2 c_{kk})$$

علماً بان (σ^2) تمثل تباين الخطأ للمجتمع والتي تعد قيمة مجهولة يمكن تقديرها بالاعتماد على بيانات العينة المختارة وعلى وفق الآتي:

$$\sigma^2 = \frac{\Sigma e^2}{n-k-1}$$

وبصيغة المصفوفات:

$$\sigma^2 = \frac{e'e}{n-k-1} \qquad \ldots \qquad (15\text{-}4)$$

وحيث ان درجات الحرية هي (n-k-1)

وسيتم برهان ذلك في الفقرة القادمة.

(7-4)القيمة التقديرية لتباين الخطأ Estimated value of Error Variance

حيث ان قيمة u الحقيقية لا يمكن معرفتها لذا فيتم تقديرها خلال استخدام العينة وبذلك فان تقدير تباين الخطأ σ^2 سيعتمد على مجموع مربعات البواقي $(e'e)$. والسؤال هو كم هي درجات الحرية التي تجعل القيمة المقدرة لتباين الخطأ غير متحيزة. والاجابة عن ذلك تكمن في الخطوات التالية:

$$e = Y - X\hat{\beta}$$

$$= Y - X(X'X)^{-1}X'Y$$
$$= \left[I_n - X(X'X)^{-1}X'\right] Y$$
$$= \underset{n\times n}{M} Y \qquad ... \qquad (16-4)$$

حيث ان: $M = I_n - X(X'X)^{-1}X'$

وهي مصفوفة متماثلة وصماء و MX = 0 .

$$\Leftarrow \quad M'M = M$$

$$\therefore e = M(X\beta + u)$$
$$= MX\beta + Mu$$
$$= Mu$$

وحيث ان M صماء

$$\Rightarrow e'e = u'M'Mu$$
$$= u'Mu$$

وبأخذ التوقع للطرفين:

$$\mathrm{E}(e'e) = \underset{(1\times n)(n\times n)(n\times 1)}{(u'Mu)}$$

وحيث ان $u'Mu$ ثابت (scalar) ⟸ الرتبة = الأثر

Rank = trace

حيث ان tr(A) يمثل حاصل جمع عناصر القطر الرئيس للمصفوفة.

$$\rho(u'Mu) = tr(u'Mu)$$
$$= \mathrm{E}\big(tr(u'Mu)\big) \qquad = \mathrm{E}\{tr(Muu')\}$$
$$= tr\ \mathrm{E}Muu' = tr\ M\mathrm{E}uu' = tr(M\sigma^2 \mathrm{I}_n)$$

$$\mathrm{E}(e'e) = \sigma^2 tr(M) \qquad ... \qquad (17-4)$$

$$tr(M) = tr\left[\mathrm{I}_n - X(X'X)^{-1}X'\right] = tr\ \mathrm{I}_n - tr(X(X'X)^{-1}X')$$
$$= tr(\mathrm{I}_n) - tr(X'X)^{-1}X'X$$
$$= n - (k+1) = n - k - 1$$

وبالتعويض في العلاقة (17-4)

$$\mathrm{E}(e'e) = \sigma^2(n - k - 1)$$

$$\therefore \sigma^2 = \frac{E(e'e)}{n-k-1}$$

أو ان:

$$\hat{\sigma}^2 = \frac{e'e}{(n-k-1)} \qquad \cdots \qquad (18-4)$$

ولكن

$$e'e = (Y-\hat{Y})'(Y-\hat{Y})$$
$$= (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$
$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$
$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'(X'X)(X'X)^{-1}X'Y$$
$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'Y$$

$$e'e = Y'Y - \hat{\beta}'X'Y \qquad \cdots \qquad (19-4)$$
$$RSS = TSS - ESS$$

ان العلاقة (19-4) تمثل حساب مجموع مربعات الاخطاء بالاعتماد على البيانات الاصلية حيث ان

$$\Sigma Y^2 = Y'Y$$

وان

$$\begin{bmatrix} \hat{\beta}_0 & \hat{\beta}_1 . & . & . & \hat{\beta}_k \end{bmatrix} \begin{pmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \vdots \\ \Sigma X_k Y \end{pmatrix} = \hat{\beta}'X'Y$$

ويمكن حسابها باستخدام القيم كانحرافات وكالاتي:

$$e'e = RSS = \Sigma Y^2 - (\hat{\beta}_0 \Sigma Y + \hat{\beta}_1 \Sigma X_1 Y + \cdots + \hat{\beta}_k \Sigma X_k Y)$$
$$= \Sigma Y^2 - \left\{ (\bar{Y} - \hat{\beta}_1 \bar{X} \cdots - \hat{\beta}_k \bar{X}_k)\ \Sigma Y + \hat{\beta}_1 \Sigma X_1 Y + \cdots \right\}$$
$$= \left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{n} \right) - \hat{\beta}_1 \left(\Sigma X_1 Y - \frac{\Sigma X_1 \Sigma Y}{n} \right) - \cdots - \hat{\beta}_k \left(\Sigma X_k Y - \frac{\Sigma X_k \Sigma Y}{n} \right)$$
$$= \Sigma y^2 - \hat{\beta}_1 \Sigma x_1 y + \hat{\beta}_2 \Sigma x_2 y \cdots + \hat{\beta}_k \Sigma x_k y)$$
$$RSS = S_{YY} - (\hat{\beta}_1 S_{X_1Y} + \hat{\beta}_2 S_{X_2Y} + \cdots + \hat{\beta}_k S_{X_kY}) \qquad \cdots \qquad (20-4)$$

والعلاقة (20-4) تحسب مجموع مربعات الاخطاء بانها الفرق بين مجموع المربعات الكلية (TSS) وبين مجموع المربعات المشروحة من قبل علاقة الانحدار وبالصيغة المصححة (ESS) أي البيانات تحسب كانحرافات عن متوسطاتها.

كما يمكن كتابة العلاقة (20-4) بصيغة اخرى بعد ضرب حدود ESS بالمقدار

$\frac{S_{X_1X_1}}{S_{X_1X_1}}$ ، $\frac{S_{X_2X_2}}{S_{X_2X_2}}$ ، $\cdots$ ، $\frac{S_{X_kX_k}}{S_{X_kX_k}}$ على التوالي فينتج:

$$RSS = S_{YY} - \left(\hat{\beta}_1 \frac{S_{X_1Y}}{S_{X_1X_1}} S_{X_1X_1} + \hat{\beta}_2 \frac{S_{X_2Y}}{S_{X_2X_2}} S_{X_2X_2} + \cdots + \hat{\beta}_k \frac{S_{X_kY}}{S_{X_kX_k}} S_{X_kX_k} \right)$$

$$= S_{YY} - (\hat{\beta}_1^2 S_{X_1X_1} + \hat{\beta}_2^2 S_{X_2X_2} + \cdots + \hat{\beta}_k^2 S_{X_kX_k}) \quad \ldots \quad (21\text{-}4)$$

$$RSS = TSS - ESS$$

وبالتعويض في المعادلة (18-4) :

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}'X'Y}{n-k-1}$$

العلاقة (19-4) تجزأ التغيرات الكلية في المتغير Y $(Y'Y)$ الى تغيرات مشروحة $(\hat{\beta}'X'Y)$ وتغيرات للخطأ $(e'e)$.

<u>برهان بعض خواص الانحدار الحسابية بالصيغة المصفوفية.</u>

1- $\mathrm{E}(e) = 0$

البرهان:

حيث ان $e = Mu$

$$\mathrm{E}(e) = \mathrm{E}(Mu) = M\mathrm{E}(u)$$
$$= 0$$

2- $X'e = 0$

بالاعتماد على المعادلة الطبيعية:

$$X'(Y - X\hat{\beta}) = X'Y - X'X\hat{\beta}$$
$$= 0$$

3- $\hat{Y}'e = 0$

$$\hat{Y}'e = (X\hat{\beta})'e$$

حيث ان: $= \hat{\beta}'X'e$

$$X'e = 0 \Rightarrow \hat{Y}'e = 0$$

(8-4)معادلة الانحدار من خلال نقطة المتوسط $\bar{X}, \bar{Y}$ Regression equation through the mean

ان انتقال الحسابات من نقطة الأصل إلى نقطة المتوسطات يعمل على تسهيل العمليات الحسابية من اجل ايجاد معالم النموذج المدروس. اذ ان جميع المتغيرات (المتغير المعتمد Y والمتغيرات المستقلة) يتم التعامل معها كانحرافات عن متوسطاتها الحسابية.

فالنموذج:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

وبأخذ المتوسط للطرفين

$$\bar{Y} = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \ldots + \beta_k \bar{X}_k + \bar{u}$$

وبالطرح:

$$Y_i - \bar{Y} = \beta_1 (X_{1i} - \bar{X}_1) + \beta_2 (X_{2i} - \bar{X}_2) + \ldots + \beta_k (X_{ki} - \bar{X}_k) + (u_i - \bar{u})$$

وبصيغة أخرى:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

اذ ان $\bar{u} = 0$ وان $x_{ki}, x_{2i}, x_{1i}, y_i$ تمثل انحرافات المتغيرات $Y_{1i}, X_{1i}, X_{2i}, X_{ki}$ عن متوسطاتها الحسابية على التوالي.

ويمكن إعادة كتابة النموذج بالصيغة المصفوفية العامة على وفق الآتي:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} x_{11} & x_{21} \ \ldots & x_{k1} \\ x_{12} & x_{22} \ \ldots & x_{k2} \\ \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & x_{kn} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix}$$

$$\Updownarrow$$

$$y = x\beta_* + u_i \qquad \ldots \qquad (22\text{-}4)$$

y : متجه انحرافات قيم المتغير المعتمد بترتيب (n×1).

x : مصفوفة انحرافات قيم المتغيرات المستقلة بترتيب (n×k) ، ويلاحظ ان ترتيبها قد انخفض عن مصفوفة X المقاسة بالقيم الأصلية.

β_*: متجـه معلمـات النمـوذج بترتيـب (k×1) ، ونلاحـظ ان المتجـه قـد حـذف منـه العنصـر β_0 ، والـذي يستوجب تقديره خارج نطاق النموذج: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} + ... + \hat{\beta}_k \bar{X}$
وبذلك تكون المعادلات الطبيعية للعلاقة (4-22) على وفق الآتي:

$$(x'x)\beta_* = x'y \qquad . . . \qquad (23\text{-}4)$$

وعليه فان المعلمات المقدرة بموجب OLS :

$$\hat{\beta}_* = (x'x)^{-1} x'y = C_* \; x'y \qquad . . . \qquad (24\text{-}4)$$

$$x'y = \begin{pmatrix} \Sigma x_1 y \\ \Sigma x_2 y \\ \vdots \\ \Sigma x_k y \end{pmatrix} \quad , \quad x'x = \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 . . . & \Sigma x_1 x_k \\ & \Sigma x_2^2 . . . & \Sigma x_2 x_k \\ \vdots & \vdots & \vdots \\ \Sigma x_k x_1 & \Sigma x_k x_2 . . . & \Sigma x_k^2 \end{bmatrix}$$

$x'x$: هي مصفوفة فيشر للمعلومات كانحرافات عن متوسطاتها وهي بترتيب (k×k) .

C_*: هي مصفوفة معكوس فيشر وهي بترتيب (k×k)

$$C_* = \begin{bmatrix} c_{11} & c_{12} & . . . & c_{1k} \\ c_{21} & c_{22} & . . . & c_{2k} \\ \vdots & \vdots & & \vdots \\ c_{k1} & c_{k2} & . . . & c_{kk} \end{bmatrix}$$

كما ان:

$$\text{var}-\text{cov}(\hat{\beta}_*) = \hat{\sigma}^2 (x'x)^{-1} = \hat{\sigma}^2 C_*$$

وهي مصفوفة لاتتضمن تباين الحد الثابت $v(\beta_0)$ والتبـاين المشترك لـ $\hat{\beta}_0$ مع $\hat{\beta}_1, . . . , \hat{\beta}_k$ وعليه يجب اشتقاق صيغة خاصة لذلك.

وللتبسيط نفرض حالة نموذج متضمن متغيرين فقط يكون فيها النموذج

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + u_i \qquad . . . \qquad (25\text{-}4)$$

ان الحد الثابت لنموذج بمتغيرين يمكن تقديره: $\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2$

وبأخذ التوقع لطرفي المعادلة (4-25) :

$$\bar{Y} = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \ldots + \bar{u}$$

اذن:

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 - \ldots - \bar{u}$$

$$\hat{\beta}_0 - \beta_0 = -\hat{\beta}_1 \bar{X}_1 + \beta_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2 + \beta_2 \bar{X}_2 + \bar{u}$$

$$= -\bar{X}_1(\hat{\beta}_1 - \beta_1) - \bar{X}_2(\hat{\beta}_2 - \beta_2) + \bar{u} \qquad \ldots \qquad (26\text{-}4)$$

$$E(\hat{\beta}_0 - \beta_0)^2 = \bar{X}_1^2 E(\hat{\beta}_1 - \beta_1)^2 + 2\bar{X}_1 \bar{X}_2 E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) + \bar{X}_2^2 E(\hat{\beta}_2 - \beta_2)^2 + E(\bar{u})^2$$

$$= \bar{X}_1^2 \operatorname{var}(\hat{\beta}_1) + 2\bar{X}_1 \bar{X}_2 \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) + \bar{X}_2^2 \operatorname{var}(\hat{\beta}_2) + \frac{\sigma^2}{n} \qquad \ldots \qquad (27\text{-}4)$$

$$= \begin{bmatrix} \bar{X}_1 & \bar{X}_2 \end{bmatrix} \begin{bmatrix} \operatorname{var}(\hat{\beta}_1) & \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_1) & \operatorname{var}(\hat{\beta}_2) \end{bmatrix} \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} + \frac{\sigma^2}{n}$$

$$\operatorname{var}(\hat{\beta}_0) = \bar{X}'\sigma^2 (x'x)^{-1} \bar{X} + \frac{\sigma^2}{n} = \sigma^2 \left[\bar{X}'(x'x)^{-1} \bar{X} + \frac{1}{n} \right] \qquad \ldots \qquad (28\text{-}4)$$

$\bar{X}$:هو متجه متوسطات المتغيرات المستقلة.

أما التباين المشترك بين $\hat{\beta}_0$ وبين كل من $\hat{\beta}_1$ و $\hat{\beta}_2$ فيمكن اشتقاقها على وفق الآتي:

تعاد كتابة العلاقة (4-26) بالصيغة: $(\hat{\beta}_0 - \beta_0) + \bar{X}_1(\hat{\beta}_1 - \beta_1) = -\bar{X}_2 E(\hat{\beta}_2 - \beta_2) + \bar{u}$

وبتربيع الطرفين وأخذ التوقع:

$$E\left[(\hat{\beta}_0 - \beta_0) + \bar{X}_1(\hat{\beta}_1 - \beta_1)\right]^2 = E\left[- \bar{X}_2 E(\hat{\beta}_2 - \beta_2) + \bar{u}\right]^2$$

$$\operatorname{var}(\hat{\beta}_0) + \bar{X}_1^2 \operatorname{var}(\hat{\beta}_1) + 2\bar{X}_1 \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = \bar{X}_2^2 \operatorname{var}(\hat{\beta}_2) + \operatorname{var}(\bar{u}) - 2\bar{X}_2 \operatorname{cov}(\hat{\beta}_2 \bar{u})$$

$$\therefore \;\; 2\bar{X}_1 \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\operatorname{var}(\hat{\beta}_0) - \bar{X}_1^2 \operatorname{var}(\hat{\beta}_1) + \bar{X}_2^2 \operatorname{var}(\hat{\beta}_2) + \operatorname{var}(\bar{u})$$

وبالتعويض عن قيمة $\operatorname{var}(\hat{\beta}_0)$ بما يساويها من العلاقة (4-27) وتبسيط العلاقة ينتج:

$$\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X}_1 \operatorname{var}(\hat{\beta}_1) - \bar{X}_2 \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) \qquad \ldots \qquad (29\text{-}4)$$

وبالطريقة نفسها يمكن التوصل الى التباين المشترك بين $\hat{\beta}_0$ و $\hat{\beta}_2$:

$$\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_2) = -\bar{X}_2 \operatorname{var}(\hat{\beta}_2) - \bar{X}_1 \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) \qquad \ldots \qquad (30\text{-}4)$$

وبدمج العلاقتين (4-29) و (4-30) بصيغة مصفوفات يتم الحصول على:

$$\begin{pmatrix} \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_2) \end{pmatrix} = -\begin{bmatrix} \operatorname{var}(\hat{\beta}_1) & \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_1) & \operatorname{var}(\hat{\beta}_2) \end{bmatrix} \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} \quad ... \quad (31-4)$$

ويمكن تعميم العلاقة (31-4) لـ k من المتغيرات المستقلة.

$$\begin{pmatrix} \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \vdots \\ \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_k) \end{pmatrix} = -\sigma^2 \begin{bmatrix} \Sigma X_1^2 & \Sigma X_1 X_2 \ ... \ \Sigma X_1 X_k \\ \vdots & \\ \Sigma X_k X_1 & \Sigma X_k X_2 \ ... \ \Sigma X_k^2 \end{bmatrix} \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_k \end{bmatrix} \quad ... \quad (32-4)$$

$$\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_j) = -\sigma^2 (x'x)^{-1} \bar{X} \quad (33-4) \quad ...$$

أي:

$$= -\sigma^2 \, C_* \, \bar{X}$$

$\bar{X}$: متجه لمتوسطات المتغيرات المستقلة بترتيب (k×1)

<u>مثال(3-4):</u> بالعودة الى المثال(2-4) حالة ثلاثة متغيرات، وذلك باستخدام البيانات كانحرافات عن متوسطاتها.

وتحسب المتوسطات: $\bar{Y} = 4 \ , \ \bar{X}_1 = 3 \ , \ \bar{X}_2 = 5$

ثم تحسب قيم المتغيرات كانحرافات عن متوسطاتها:

جدول (3-4)

$y = Y - \bar{Y}$	$x_1 = X_1 - \bar{X}_1$	$x_2 = X_2 - \bar{X}_2$	y^2	x_1^2	x_2^2	$x_1 x_2$	$x_1 y$	$x_2 y$
-1	0	0	1	0	0	0	0	0
-3	-2	-1	9	4	1	2	6	3
4	2	1	16	4	1	2	8	4
-1	-1	-1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
Σ	0	0	28	10	4	6	16	9

$$\begin{bmatrix} x'x & \vdots & x'y \end{bmatrix} = \left[\begin{array}{cc:c} 10 & 6 & 16 \\ 6 & 4 & 9 \end{array}\right]$$

$$(x'x)^{-1}=\begin{pmatrix}1 & -\frac{3}{2}\\ -\frac{3}{2} & \frac{5}{2}\end{pmatrix}=C_* \Rightarrow \begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\begin{pmatrix}1 & -\frac{3}{2}\\ -\frac{3}{2} & \frac{5}{2}\end{pmatrix}\begin{pmatrix}16\\ 9\end{pmatrix}=\begin{pmatrix}2.5\\ -1.5\end{pmatrix}$$

ثم نستخرج قيمة $\hat{\beta}_0$ على وفق القانون:

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}_1-\hat{\beta}_2\overline{X}_2$$

$$=4-2.5(3)-(-1.5(5))$$

$$=4$$

اذن معادلة الانحدار:

$$\hat{Y}=4+2.5X_1-1.5X_2$$

اما تباين المعلمات المقدرة

$$\text{var}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\sigma^2 C_*$$

$$\text{var}(\hat{\beta}_1)=\sigma^2 c_{11}=\sigma^2(1)$$

$$\text{var}(\hat{\beta}_2)=\sigma^2 c_{22}=\sigma^2(5/2)$$

وبتطبيق العلاقة (4-28) يمكن الحصول على:

$$\text{var}(\hat{\beta}_0)=\sigma^2\left[\overline{X}'(x'x)^{-1}\overline{X}+\frac{1}{n}\right]=\sigma^2\left[\begin{pmatrix}3 & 5\end{pmatrix}\begin{bmatrix}1 & -\frac{3}{2}\\ -\frac{3}{2} & \frac{5}{2}\end{bmatrix}\begin{pmatrix}3\\ 5\end{pmatrix}+\frac{1}{5}\right]$$

$$\text{var}(\hat{\beta}_0)=\sigma^2(156+\frac{1}{5})$$

أما قيمة $\hat{\sigma}$ فيمكن تقديرها على وفق:

$$\frac{RSS}{n-k-1}=\frac{RSS}{5-3}=\frac{RSS}{2}$$

أما

$$RSS=TSS-ESS$$

$$TSS=28$$

$$ESS=\hat{\beta}_*'\;x'y$$

$$= (2.5 \quad -1.5) \begin{pmatrix} 16 \\ 9 \end{pmatrix}$$

$$= 40 - 13.5 = 26.5$$

$$RSS = 28 - 26.5 = 1.5$$

$$\Rightarrow \quad \hat{\sigma}^2 = 0.75$$

(9-4) معامل التحديد R^2 بالصيغة المصفوفية. Coefficient of Determination using matrix form

معامل التحديد وكما تم تعريفه في الفصل الثالث :

$$R^2 = \frac{ESS}{TSS}$$

ففي حالة الانحدار البسيط:

$$R^2 = \frac{\hat{\beta}_1 \Sigma xy}{\Sigma y^2} \quad , \quad k = 1$$

في الانحدار المتعدد:

$$R^2 = \frac{\hat{\beta}_1 \Sigma x_1 y + \hat{\beta}_2 \Sigma x_2 y}{\Sigma y^2} \quad , \quad k = 2$$

أما عندما يكون عدد المتغيرات = k

$$R^2 = \frac{\hat{\beta}_1 \Sigma x_1 y + \hat{\beta}_2 \Sigma x_2 y + \ldots + \hat{\beta}_k \Sigma x_k y}{\Sigma y^2}$$

وبالصيغة المصفوفية:

$$R^2 = \frac{\hat{\beta}' x' y}{y' y} \quad \ldots \quad (34\text{-}4)$$

معامل الارتباط المتعدد Multiple correlation coeffecient

ويرمز له بالرمز R

وهو مقياس للعلاقة بين Y و $\hat{Y}$. $(R_{Y\hat{Y}})$ ويعبر عنه بالقانون:

$$R_{Y\hat{Y}} = \frac{\Sigma(Y - \bar{Y})(\hat{Y} - \bar{Y})}{\sqrt{\Sigma(Y - \bar{Y})^2 \Sigma(\hat{Y} - \bar{Y})^2}}$$

وهو يمثل الجذر ألتربيعي لمعامل التحديد.

معامل التحديد المعدل R^2Adjusted

ويرمز له $\bar{R}^2$ ، ويستخدم هذا المؤشر في الانحدار المتعدد لانه يعطي دلالة أوضح من R^2 حول جودة (حسن ملاءمة) النموذج.

حيث ان R^2 تزداد، عادة، بإضافة متغيرات مستقلة جديدة في النموذج (بغض النظر عن مدى ملاءمتها له). وتكمن الصعوبة مع R^2 هو عدم وجود تحديد على زيادة عدد المتغيرات المستقلة المستخدمة في تفسير المتغير التابع. في حين ان معامل التحديد المعدل $\overline{R}^2$ يراعي نسبة الانخفاض في تباين (او تغيير) المتغير Y والتي تعزا لإضافة X_i للنموذج الذي يحوي X_j . ويكون معامل التحديد المعدل:

$$\overline{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}$$

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2) \quad \ldots \quad (35-4)$$

ويمكن ان تزداد قيمة $\overline{R}^2$ او تظل على حالها وقد تنقص اذا أضفنا متغيرات مستقلة إضافية للنموذج. وذلك لان مثل هذه المتغيرات المستقلة الإضافية سوف تخفض عادة من قيمة RSS ولكنها وبالوقت نفسه سوف تقلل من (n-k-1) وبعبارة أخرى يفضل عدم إضافة متغير للنموذج اذا تسببت إضافته الى تخفيض قيمة $(\overline{R}^2)$ ويتضح من العلاقة (35-4) ان: $\overline{R}^2 \langle R^2$

مثال(4-4): باستخدام بيانات الجدول (2-4) نفسها فان:

$$TSS = \Sigma y^2 = 28$$

$$ESS = \hat{\beta}'x'y = \begin{bmatrix} 2.5 & -1.5 \end{bmatrix} \begin{bmatrix} 16 \\ 9 \end{bmatrix} = 26.5$$

وبذلك فان مجموع مربعات الخطأ:

$$RSS = 28 - 26.5 = 1.5$$

وعليه يكون معامل التحديد للنموذج (معامل الارتباط المتعدد)

$$R^2_{Y \cdot X_1 X_2} = \frac{26.5}{28} = 0.9464$$

أي ان استخدام المتغيرين X_1 و X_2 تمكنت من تفسير 94.64% من التغيرات الإجمالية في Y.

أما معامل التحديد المعدل:

$$\overline{R}^2 = 1 - \frac{4}{2}(1 - 0.9464) = 0.8928 \langle R^2$$

(10-4) <u>الانحدار المتعدد والانحدار البسيط</u>.Multiple VS simple regression

هل توجد علاقة بين معاملات الانحدار المتعدد ومعاملات البسيط ؟ للإجابة عن هذا السؤال، لنمعن النظر في علاقات الانحدار التالية:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

$$Y = a_0 + b_1 X_1 \qquad + v_1$$

$$Y = c_0 + b_2 X_2 \qquad + v_2$$

$$X_2 = d_0 + b_{21} X_1 \qquad + v_3$$

$$X_1 = e_0 + b_{12} X_2 \qquad + v_3$$

حيث ان:

$$\beta_1 = \frac{\partial Y}{\partial X_1}\bigg| \begin{pmatrix} \text{استبعاد أثر } X_1 \\ X_1 \text{ constant} \end{pmatrix} \qquad \beta_2 = \frac{\partial Y}{\partial X_2}\bigg| \begin{pmatrix} \text{استبعاد أثر } X_2 \\ X_2 \text{ constant} \end{pmatrix}$$

$$b_1 = \frac{dY}{dX_1} \qquad , \qquad b_2 = \frac{dY}{dX_2}$$

$$b_{21} = \frac{dX_2}{dX_1} \qquad , \qquad b_{12} = \frac{dX_1}{dX_2}$$

ويمكن اثبات العلاقات التالية:

$$\beta_1 = \frac{b_1 - b_2 b_{21}}{1 - R_{21}^2} \qquad . . . \qquad (36 - 4)$$

وكذلك:

$$\beta_2 = \frac{b_2 - b_1 b_{12}}{1 - R_{12}^2} \qquad . . . \qquad (37 - 4)$$

ومن تفحص المعادلات (4-36) و (4-37) يتضح انه في حال انعدام العلاقة بين X_1 و X_2 فان b_{21} و b_{12} تساوي صفراً وكذلك R_{12}^2 و R_{21}^2 تساوي صفراً .

وبذلك: $\beta_1 = b_1$ وكذلك $\beta_2 = b_2$

أي ان معاملات الانحدار الجزئية في الانحدار المتعدد = معاملات الانحدار المقابلة لها في الانحدار البسيط اذا كان الارتباط بين المتغيرات التوضيحية معدوماً.

- وبشكل عام اذا كانت مصفوفة المعلومات مصفوفة فيشر $(X'X)$ مصفوفة قطرية فان المعلمات المقدرة من الانحدار المتعدد تكافئ معلمات الانحدار المقابلة لها في الانحدار البسيط.

برهان هذه الحالة تترك كتمرين.

(11-4)الأثر المباشر والأثر غير المباشر Direct & Indirect effect.

ان الافتراض الذي ينص على تشخيص النموذج بشكل صحيح فضلاً عن كونه خالياً من الأخطاء تعد فرضية مهمة. وعلى أساس تحقق هذه الفرضية فان المقدرات تتمتع بصفة (BLUE).

لتوضيح فكرة هذا المبحث، نفترض ان النموذج الصحيح هو:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_t \;\; . \; . \; . \qquad (38\text{-}4)$$

وبذلك فان معلمات الانحدار المتعدد المقدرة $\hat{\beta}_1$, $\hat{\beta}_2$ تمثل المشتقات الجزئية:

$$\hat{\beta}_1 = \frac{\partial Y}{\partial X_1} \quad \& \quad \hat{\beta}_2 = \frac{\partial Y}{\partial X_2}$$

ومع تحقق الفروض فان:

$$\mathrm{E}(\hat{\beta}_1) = \beta_1$$

$$\& \quad \mathrm{E}(\hat{\beta}_2) = \beta_2$$

اذا تم الافتراض ان الباحث اعتمد نموذج الانحدار البسيط في التقدير كالآتي:

$$Y = b_0 + b_{01} X_1 + u_t \;\; . \; . \; . \qquad (39\text{-}4)$$

فان $b_{01} = \dfrac{dY}{dX_1}$ يمثل أثر المتغير X_1 على Y.

فالسؤال المطروح: هل ان المعلمة (b_{01}) ستوفر تقدير غير متحيز لـ β_1. بمعنى $\mathrm{E}(b_{01}) = \beta_1$

علماً بان المتغير X_2 تم حذفه من المعادلة(- 38)4.

الجواب: بشكل عام $\mathrm{E}(b_{01}) \neq \beta_1$

حيث يمكن البرهنة على ان:

$$b_{01} = \beta_1 + \beta_2 b_{21} X_1 + error$$

اذ ان: (b_{21}) هي معلمة انحدار X_2 (المحذوف) على X_1، أي:

$$X_{2t} = c_2 + b_{21} X_1 + error$$

$$\hat{b}_{21} = \frac{\Sigma x_2 x_1}{\Sigma x_1^2}$$

أي ان: $\mathrm{E}(b_{01}) = \beta_1 + b_{21} \cdot \beta_2$

وبشكل عام: $b_{21} \cdot \beta_2 \neq 0$

اذن (b_{01}) متحيزة.

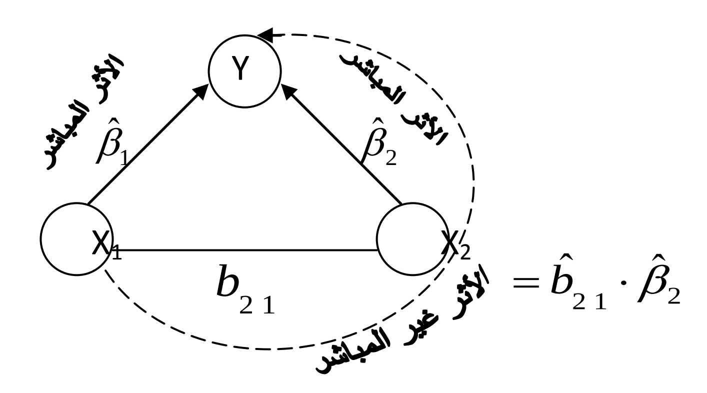
التقديم السابق له تفسير مهم في توضيح فكرة الأثر المباشر أو الصافي لـ X_1 على Y. وذلك من خلال جعل أثر X_2 ساكناً. إلى جانب قياس الأثر غير المباشر على Y نسبة إلى أثره على المتغير المحذوف X_2.

وبذلك فان β_1 يقيس فقط الاثر المباشر او الصافي لـ X_1 على Y. في حين (b_{01}) يقيس الأثر الإجمالي لـ X_1 على Y ويبقى الاثر غير المباشر لـ X1 على Y يكون ممثلاً بـ $(\hat{\beta}_2 \cdot \hat{b}_{21})$ عبر أثر X_2.

ويمكن توضيح ذلك من خلال المخطط التالي:

مخطط (4-1)

الأثر المباشر والأثر غير المباشر في نموذج ثلاثة متغيرات.



مثال (4-6):

$$Y' = \begin{bmatrix} 1 & 3 & 8 \end{bmatrix} \quad , \quad X_1' = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \quad , X_2' = \begin{bmatrix} 2 & 1 & -3 \end{bmatrix}$$

قدر العلاقات التالية:

$$(i)\ Y_i = \alpha_0 + \alpha_1 X_{1i} + u_{1i}$$

$$(ii)\ Y_i = \lambda_0 + \lambda_2 X_{2i} + u_{2i}$$

$$(iii)\ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_{it}$$

هل ان:

(أ) $\hat{\alpha}_1 = \hat{\beta}_1$ ولماذا ؟

(ب) $\hat{\lambda}_2 = \hat{\beta}_2$ ولماذا ؟

الحل:

$$\Sigma X_1 = 6 \ , \ \Sigma X_2 = 0 \ , \ \Sigma Y = 12 \ , \ \Sigma X_1 Y = 31 \ , \ \Sigma X_2 Y = -19$$

$$\Sigma X_1^2 = 14 \ , \ \Sigma X_2^2 = 14 \ , \ \overline{X}_1 = 2 \ , \ \overline{X}_2 = 0 \ , \ \overline{Y} = 4$$

$$\hat{\alpha}_1 = \frac{\Sigma X_1 Y - \frac{\Sigma X_1 \Sigma Y}{n}}{\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n}} = \frac{31 - \frac{(6)(12)}{3}}{14 - \frac{6^2}{3}}$$

$$= \frac{31 - 24}{14 - 12} = \frac{7}{2} = 3.5$$

$$\hat{\lambda}_1 = -1.286$$

$$y' = \begin{bmatrix} -3 & -1 & 4 \end{bmatrix} \ , \ x_1' = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \ , \ x_2' = \begin{bmatrix} 2 & 1 & -3 \end{bmatrix}$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (x'x)^{-1}x'y = \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 \\ & \Sigma x_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma x_1 y \\ \Sigma x_2 y \end{bmatrix} = \begin{bmatrix} 2 & -5 \\ & 14 \end{bmatrix}^{-1} \begin{bmatrix} 7 \\ -19 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{14}{3} & \frac{5}{3} \\ \frac{5}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 7 \\ -19 \end{bmatrix} = \begin{bmatrix} 32.67 - 31.67 \\ 11.67 - 12.67 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \hat{\beta}_1 = 1 \ , \ \hat{\beta}_2 = -1$$

$\hat{\beta}_1 = 1$ في حين $\hat{\alpha}_1 = 3 = b_{0\,1}$

وكذلك

$\hat{\beta}_2 = -1$ $\hat{\lambda}_2 = -\frac{9}{7}$

$\hat{\beta}_1 \neq \hat{\alpha}_1$ لان $\hat{\beta}_1$ يمثل الأثر المباشر لـ X_1 علىY.

في حين $\hat{\alpha}_1$ يمثل الأثر الإجمالي وهو يمثل الأثر المباشر+ الأثر غير المباشر من خلال المتغيرX_2.

والفرق يكمن في الأثر غير المباشر. $b_{01} = (\hat{\alpha}_1 - \hat{\beta}_1 + \hat{\beta}_2 \hat{b}_{2\,1})$

وكذلك $\hat{\beta}_2 \neq \hat{\lambda}_2$

حيث ان $\hat{\beta}_2$ تمثل الأثر المباشر فقط لـ X_2 على Y.

في حين ان $\hat{\lambda}_2$ تمثل الأثر الصافي (الإجمالي) والذي يمثل الأثر المباشر + الأثر غير المباشر من خلال المتغير X_1 ، وذلك بتطبيق العلاقة: $X_1 = b_1 + b_{12} X_2 + error$

والاختلاف يكمن في الأثر غير المباشر $(\hat{\beta}_1 \cdot \hat{b}_{1\,2})$ أي: $\hat{\lambda}_2 = \hat{\beta}_2 + \hat{b}_{12} \hat{\beta}_1$

<u>مثال (4-7):</u> عينة بحجم (13) مشاهدة لقيم Y و X_1 و X_2 ، أعطت النتائج التالية:

(1) $\hat{Y} = 7.2 - 14X_1 + 1.5X_2 \qquad r^2 = 0.87$

(2) $\hat{Y} = 6.12 + 0.25X_1 \qquad r^2 = 0.013$

$t:$ (1.4) (0.3)

$$(3)\ \ X_2 = -0.7 + 1.1X_1 \qquad\qquad r = 0.4$$

$$t: \quad (-0.2)\ \ (2.7)$$

فسر نتائج التقدير؟

الحل: $\hat{\beta}_1 = -1.4$ تمثل الأثر المباشر لـ X_1 على Y.

$\hat{b}_{0\,1} = 0.25$ تمثل الأثر الكلي (الصافي) لـ X_1 على Y.

$\hat{b}_{2\,1} = 1.1$ تمثل الأثر الصافي لـ X_1 على X_2.

$\hat{\beta}_2 = 1.5$ تمثل الأثر المباشر لـ X_2 على Y.

الأثر غير المباشر لـ X_1 على Y عبر $X_2 \equiv (\hat{\beta}_2 \cdot b_{2\,1} = (1.1)(1.5) = 1.65)$

فالمعادلة (3) تعني عند زيادة X_1 بمقدار وحدة واحدة فان X_2 يزداد بمقدار (1.1) . ولكن عند زيادة X_2 بهذا القدر فان أثره في Y سيكون $(\hat{\beta}_2 \cdot b_{2\,1}) = (1.5)\ (1.1) = 1.65$.

وعليه فان الأثر الصافي $1.65 + (-1.4) =$

$=0.25$

(12-4)<u>معامل الارتباط الجزئي (Partial correlation coefficient)</u>

في حالة وجود متغيرين، فان معامل الارتباط البسيط يعني درجة الترابط الخطي بين المتغيرين Y و X . وعندما ندرس حالة ثلاثة متغيرات وأكثر، لابد من إعطاء عناية خاصة لتفسير معامل الارتباط بين أي متغيرين. في حالة ثلاثة متغيرات Y و X_2 و X_3 يمكن حساب معاملات الارتباط التالية: r_{12} معامل ارتباط Y مع X_2 ، ومعامل ارتباط Y مع $X_3(r_{13})$ و r_{23} معامل ارتباط X_2 و X_3. ويطلق عليها معاملات الارتباط البسيط او معامل ارتباط من الدرجة صفر (correlation coefficient of zero order) ، والتي يمكن حسابها بموجب العلاقة :

$$r_{ij} = \frac{\Sigma x_i x_j}{\sqrt{\Sigma x_i^2}\sqrt{\Sigma x_j^2}}$$

والسؤال المطروح هنا، هل ان (r_{12}) سوف يعكس درجة الترابط بين المتغيرين Y و X_2 في حالة وجود المتغير X_3 والذي قد يكون مرتبطاً مع كليهما. والإجابة ان r_{12} بشكل عام لا يعكس درجة الترابط بين المتغيرينY و X_2 في حال وجود متغير ثالث (X_3) وذلك لان قسماً من التغيرات في Y هي بسبب المتغير X_3. وفي حقيقة الأمر فان r12 يعكس انطباعاً خاطئاً لطبيعة الترابط بين Y و X_2 ويتطلب الأمر معامل ارتباط آخر يسمى معامل الارتباط الجزئي $(r_{12.3})$ و $(r_{13.2})$ و $(r_{23.1})$.

$r_{12.3}$:معامل الارتباط بين Y و X_2 بافتراض ان X_3 ثابت.

$r_{13.2}$: معامل الارتباط بين Y و X_3 بافتراض ان X_2 ثابت.

$r_{23.1}$: معامل الارتباط بين X_2و X_3 بافتراض ان Y ثابتاً.

وتحسب القيم على وفق المعادلات التالية:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13}^2}}$$

ويمكن تعميم ذلك لاي ثلاثة متغيرات k , j , i فان معامل الارتباط الجزئي يتبع الصيغة التالية:

$$r_{ij.k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{1 - r_{ik}^2}\sqrt{1 - r_{jk}^2}} \quad \ldots \quad (40-4)$$

والعلاقة (4-40) تسمى معامل الارتباط الجزئي للدرجة الأولى، وهكذا $r_{12.34}$هو معامل الارتباط الجزئي للدرجة الثانية و $r_{12.345}$ هو معامل الارتباط للدرجة الثالثة، في حين r_{12} و r_{13} و r_{23} هو معامل الارتباط من الدرجة صفر. ويتضح من العلاقة (4-40) ان معامل الارتباط الجزئي يكون بدلالة معاملات الارتباط البسيط (من الدرجة صفر).

ان معامل الارتباط الجزئي $r_{12.3}$ يمكن حسابه بإجراء معامل الارتباط بين البواقي الناتجة من انحدار (Y على X_3) وبين بواقي انحدار (X_2 على X_3) وبالطريقة نفسها يمكن استنتاج $r_{13.2}$ و $r_{23.1}$ وبشكل عام فان معامل الارتباط الجزئي بين متغيرين يقاس بعلاقة الارتباط بين البواقي لكل متغير بعد إزالة الأثر الخطي لجميع المتغيرات التوضيحية التي يتم تثبيتها.

وبالاعتماد على العلاقة (4-40) يمكن التوصل إلى الاستنتاجات التالية:

(١) اذا $r_{ij} = 0$ فان $r_{ij.k}$ تبقى مختلفة عن الصفر ما لم يتحقق أحد الشرطين او كلاهما:
أما $r_{ik} = 0$ أو $r_{jk} = 0$ أو كلاهما.

(٢) اذا $r_{ij} = 0$ و $r_{jk} \neq 0$ وبالإشارة نفسها فان $r_{ij.k}$ سالبة.
في حين تكون موجبة اذا اختلفت r_{ik} و r_{jk} بالإشارة.

(٣) ان إشارة (r_{ij}) و ($r_{ij.k}$) ليست بالضرورة متطابقة.

(٤) اذا $r_{ik} = 0$ و $r_{jk} = 0$ فان r_{ij} ليس بالضرورة يساوي صفراً.

(٥) $0 \le r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}\ r_{13}\ r_{23} \le 1$

(٦) في حالة ثلاث متغيرات، ان التغيرات في المتغير (i) يمكن تجزئتها إلى جزأين، الأول يمثل التغيرات في المتغير (i) والتي يتم توضيحها من قبل المتغير (j) فقط (r_{ij}^2). أما الجزء الآخر فهو الجزء غير الموضح من قبل المتغير (j)وهو $(1-r_{ij}^2)$ومرجحاً بمقدار الجزء الموضح بواسطة المتغير (k) بعد جعل تأثير المتغير (j) ثابتاً:

$$R_{i.jk}^2 = r_{ij}^2 + (1-r_{ij}^2)\ r_{ik.j}^2 \quad ... \quad (41-4)$$

$R_{i.jk}^2$ معامل التحديد لانحدار المتغير (i) على المتغيرين (j) و (k).

أو:

$$R_{i.jk}^2 = r_{ik}^2 + (1-r_{ik}^2)\ r_{ij.k}^2 \quad ... \quad (41-4)$$

(٧) وبشكل عام: معامل التحديد في نموذج الانحدار المتعدد يكون اكبر من معامل التحديد في نموذج الانحدار البسيط اذا كان معامل التحديد الجزئي موجباً.

أي:
$$R_{i.jk}^2 \rangle\ r_{ij}^2 \quad iff \quad r_{ik.j} \rangle\ 0$$

(٨) وبإعادة صياغة العلاقة (41-4) يتم الحصول على:

$$r_{ij.k}^2 = \frac{R_{i.jk}^2 - r_{ik}^2}{1-r_{ik}^2} \quad ... \quad (42-4)$$

حيث ان البسط في العلاقة (42-4) يمثل مجموع المربعات المشروحة من قبل المتغير (j) فقط بعد ازالة أثر المتغير (k) . أي: $ESS(j/k)$

أما المقام $(1-r_{ik}^2)$ فهو يمثل مجموع مربعات الخطأ من انحدار i على المتغير (k) أي: $RSS(k)$.

(٩) مربع معامل الارتباط الجزئي $(r_{ij.k}^2)$ يفسر مربع التغيرات في المتغير (i) والتي يتم توضيحها من قبل المتغير (j) في حين يتم عزل أثر المتغير (k) . وهو بذلك مماثل لمعامل التحديد، فهو يقيس الأهمية النسبية للتغيرات المشروحة من قبل المتغير (j) باعتبار أثر المتغير k يبقى ثابتاً وهو بذلك ينفع لقياس جودة المواءمة الجزئية اذا كان ترميز المتغيرات على وفق التي:

المتغير i : هو المتغير المعتمد (التابع).

المتغير j ، والمتغير k هما المتغيران المستقلان.

وعليه فان العلاقة (4-42) تعكس الحقيقة التالية:

$$r^2_{ij.k} = \frac{ESS(j/k)}{RSS(k)} \qquad \ldots \qquad (43\text{-}4)$$

ويمكن تفسيره بانـه نسبة الانخفاض في تباين المتغير المعتمد (i) والتي تعزا لإضافة المتغير (j) للنموذج والذي يتضمن المتغير التوضيحي (k).

وفي حالة أربعة متغيرات (متغير معتمد i(Y) وثلاثة متغيرات توضيحية j ، k ، L وهي X_2 ، X_3 ، X_4.

فان نسبة الانخفاض في تباين Y بسبب إضافة المتغير X_2 للنموذج الذي يتضمن X_3 و X_4 وهي:

$$r^2_{YX_2.X_3X_4} = \frac{ESS(X_2/X_3X_4)}{RSS(X_3X_4)}$$

علاقة معامل التحديد وتباين معلمة الانحدار : Relation between coefficient of determination and regression parameter

في نموذج الانحدار $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$ يمكن تحديد قيمة تباين معلمتي الانحدار $\hat{\beta}_1$ و $\hat{\beta}_2$ بدلالة مربع معامل الارتباط كالآتي:

$$\operatorname{var}(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma x_1^2} \cdot \frac{1}{(1-r_{12}^2)} \qquad \& \qquad \operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_2^2} \cdot \frac{1}{(1-r_{12}^2)}$$

حيث ان r_{12}^2 هو معامل التحديد لانحدار X_1(على X_2) أو لانحدار X_2 على X_1 .

وبشكل عام في الانحدار المتعدد: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ldots \ + \beta_k X_k + u$

فان:

$$\operatorname{var}(\hat{\beta}_j) = \frac{1}{\Sigma x_j^2} \cdot \frac{1}{(1-R_j^2)} \qquad \ldots \qquad (44\text{-}4)$$

$$; \forall \quad j = 1, \ \ldots \ , k$$

حيث ان R_j^2 يمثل معامل التحديد لانحدار المتغير X_j على جميع المتغيرات التوضيحية:

$$X_j = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ \ldots \ + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \ \ldots \ + \alpha_k X_k + e$$

أسئلة الفصل الرابع:

س1: في نموذج خطي بثلاثة متغيرات Y، X_1، X_2 حدد المعادلات الطبيعية لتقدير معادلة الانحدار من خلال نقطة المتوسط ومن خلال نقطة الاصل.

س2: فسر دلالة الفرضيات التالية:

١. رتبة المصفوفة $X \left[\rho(X)\right]$ تامة من ناحية الأعمدة.

٢. u متغير عشوائي كروي .spherical dist

٣. المصفوفةX مصفوفة لمشاهدات المتغيرات المستقلة.

س3: أ. ما هو المعيار الذي تستند إليه طريقة المربعات الصغرى الاعتيادية OLS.

اذكر معايير اخرى تستخدم في التقدير.

ب. يعتمد مبدأ المربعات الصغرى على جملة من الفرضيات ، ناقشها من الناحية العملية.

س4: اذا توافرت المعلومات التالية:

$$n = 20; \quad \bar{Y} = 20; \quad \bar{X}_2 = 5; \quad \bar{X}_1 = 10; \quad y'y = 60; \quad x'y = \begin{bmatrix} 20 \\ 16 \end{bmatrix}; \quad x'x = \begin{bmatrix} 24 & 8 \\ & 20 \end{bmatrix}$$

1. قدر معلمات النموذج وفسر دلالاتها الاحصائية.

٢. الانحراف المعياري للمقدرات.

٣. تنبأ عن قيم y عندما $x_2 = 9 \quad , \quad x_1 = 1$

س٥: في نموذج الانحدار الخطي المتعدد كيف تعبر إحصائياً عن عدم وجود علاقة خطية تامة بين المتغيرات التوضيحية .

س6: : لوحة بيانات انحدار Y علىX_1، X_2 ، أعطت معامل التحديد $R^2 = 0.79$ علما ان البيانات هي:

Y	٨	٩	١٣	٢٠	٢١
X_1	١	٢	٣	٢.٥	٤
X_2	٥	٧	٩	١٠	١١

المطلوب :

١. فسر دلالة R^2 للمعادلة.

٢. احسب قيمته المعدلة.

٣. احسب تباين الخطأ لمعادلة التقدير.

س7: مع توافر لوحة المعلومات التالية:

$$TSS = 1260.89 \ , \ n = 20 \ , \ ESS(X_2) = 1175.19$$

اختبر الاثر الاضافي لتضمينX_3 لمعادلة الانحدار : $Y = \beta_0 + \beta_1 X_2 + U$

س8: اختبر معنوية تساوي معلمتي الانحدار (مستوى دلالة ٥%) اذا علمت : $r_{X_1X_2} = 0.5$

$$\hat{Y} = 2.7 + 2.1X_1 - 0.78X_2$$
$$t: \quad (3.7) \quad (-2.5)$$

س9: المتغير المعتمد Y_i يرتبط بالمتغيرات المستقلة X_{ij} على وفق النموذج الخطي العام التالي:

$Y = X\beta + U$ حيث : $U \sim NID(0, \sigma^2 I_n)$. اشتق الصيغة لتقدير متجه المعلمات (β) باستخدام طريقة الامكان الاعظم (ML) .

ثم احسب قيمة المقدرات من واقع البيانات:

$$(x'x)^{-1} = \begin{bmatrix} 0.86 & -0.15 & -0.07 \\ & 32 & -0.45 \\ & & 0.55 \end{bmatrix} \quad , \quad x'y = \begin{bmatrix} 9 \\ 12 \\ 16 \end{bmatrix}$$

س10 : في النموذج: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$

اشتق صيغة ملائمة مصفوفة التباين والتباين المشترك للمعلمة β_0 المقدرة .

س11: البيانات التالية تخص عدد الوحدات المصنعة (X) وكلفة العمل الكلية المرافقة لها (Y) لـ ١٢ منشأة متماثلة.

X	35	25	40	35	64	25	50	67	69	50	70	50
Y	112	125	128	115	162	130	142	158	175	140	170	145

هل ان النموذج الخطي مطابق للبيانات . استخدم مستوى دلالة ٥%.

س12: حدد صحة العبارات التالية موضحا جوانب الخطأ في حالة كون العبارة خاطئة:

١. حسن التطابق يقيس تشتت المشاهدات عن خط الانحدار المقدر ويقاس بمعيار متوسط مربعات الخطا.

٢. ان فرضية استقلال المتغيرات التوضيحية $X's$في معادلة الانحدار تعني ان المتغيرات $X's$ ثابتة للعينة المختارة.

٣. ان تباين المعلمات المقدرة بطريقة OLS تتبع القانون $\text{var}(\hat{\beta}_i) = \dfrac{\sum x^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$ بشكل عام.

٤. عند انحدار Y على X_1 ، X_2فان $r^2_{Y1.2}$ تمثل النسبة التي تم توضيحها من قبل X_1 لتباين Y المتبقي دون ان يوضحه X_2.

س13: مع توافر لوحة البيانات:

Y	3	5	6	10	10	11
X_1	1	2	1	3	٣	4
X_2	3	2	3	3	2	3

١- احسب معاملات الارتباط البسيط.

٢- احسب معاملات الارتباط الجزئي.

٣- اذكر خطوات حسابها بطريقة الانحدار.

الفصل الخامس
الاستدلال في نموذج الانحدار المتعدد
Inference in multiple linear regression

يعد هذا الفصل استمراراً للفصل الثالث. فهو توسيع لفكرة الاستدلال في نموذج المربعات الصغرى عندما يكون عدد المتغيرات التوضيحية متغيرين أو أكثر ومع التركيز على بعض التمايز عن الانحدار البسيط والتطرق إلى الأنواع المختلفة من الاختبارات في الانحدار المتعدد.

في هذا الصدد لابد من التأكيد على فرضية التوزيع الاحتمالي للمتغير العشوائي (u)والذي تم افتراضه بأنه يتبع التوزيع الاحتمالي الطبيعي (Normal) بمتوسط صفري وتباين ثابت σ^2 حيث إن هذا الافتراض مع الافتراضات الأخرى وبتطبيق المربعات الصغرى فان المعلمات المقدرة تتميز بصفة (BLUE) ، فضلاً عن إنها تتبع التوزيع الطبيعي ويمكن تحديد التوزيع للمعلمات المقدرة:

$$\hat{\beta}_j \sim N(\beta_j,\ \sigma^2 c_{jj}) \quad ، j = 0,1,2, . . . , k$$

وبتعويض قيمة σ^2 بالقيمة المقدرة $\hat{\sigma}^2$ فان:

$$\frac{\hat{\beta}_j - \beta_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t_{(n-k-1)} \qquad . . . \qquad (1\text{-}5)$$

وبذلك فان توزيع ستيودنت t يمكن استخدامه لتحديد مجالات الثقة فضلاً عن اختبارات المعنوية حول معلمات الانحدار الجزئية الحقيقية. وفي هذا الصدد سيتم عرض فقرات هذا الفصل على وفق الآتي:

(5−1)اختبار الفرضيات حول معلمة الانحدار الجزئية. Testing hypothesis about partial regression estimates

لتوضيح آلية الاختبار، فان فرضية العدم:

(أ) $H_0: \beta_j = 0$ vs. $H_1: \beta_j \neq 0 \ \forall \ j = 1,2,\cdots,k$

(ب) وبشكل عام : $H_0: \beta_j = \beta_{j0}$ vs. $H_1: \beta_j \neq \beta_{j0}$

حيث ان β_{j0}: تمثل قيمة المعلمة عند نقطة محددة (معلومة).

ولاختبار هذه الفرضية يتم استخدام المختبر t في العلاقة (5−1) ، ثم تقارن القيمة المحسوبة لـ t بالقيمة الجدولية وبدرجات حرية (n−k−1)، ولمستوى معنوية محددة (α)، $t_{c(n-k-1,\frac{\alpha}{2})}$ ، فيكون القرار برفض فرضية العدم. إذا $t^*_{\hat{\beta}_j} \ \rangle \ t_c$ ، أي إن المعلمة المقدرة $\hat{\beta}_j$ تختلف معنوياً عن الصفر في (أ)، أو تختلف معنوياً عن القيمة المعلومة β_j. في الحالة (ب).

ولابد من تحديد التقابل بين اختبارات المعنوية حول المعلمات الجزئية ومجال الثقة المقدر فان $(1-\alpha)\%$ ثقة للمعلمة β_j هو:

$$\hat{\beta}_j - t_{c(\frac{\alpha}{2})} \cdot s.e(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t_{c(\frac{\alpha}{2})} \cdot s.e(\hat{\beta}_j)$$

أي:

$$L \le \beta_j \le U$$

بمعنى أن (β_j) تقع ما بين الحدين الأدنى L والحد الأعلى U وباحتمال ثقة $(1-\alpha)\%$

بعبارة أخرى :إذا تم سحب (100) عينة كل بحجم (n) من المشاهدات وتم عمل (100) فترة ثقة للمعلمة β_j $(\hat{\beta}_j \pm s.e(\hat{\beta}_j) \cdot t_{c(\frac{\alpha}{2})})$ فأننا نتوقع 95% من هذه الفترات سوف تتضمن معلمة المجتمع (β_j).

مثال:(1−5) اعتماداً على المشاهدات في المثال (2−4) فان معادلة التقدير:

$$\hat{Y} = 4 + 2.5X_1 - 1.5X_2$$

ولاختبار معنوية معلمتي الانحدار الجزئية نتبع الآتي:

لاختبار معنوية المعلمة β_1 .

- الفرضية: $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \ne 0$
- نحسب النسبة t لفرضية العدم:

$$t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - 0}{s.e(\hat{\beta}_1)}$$

وبما إن:

حيث إن: $s.e(\hat{\beta}_1) = \hat{\sigma}\sqrt{c_{11}}$

$$(x'x) = \begin{pmatrix} 10 & 6 \\ 6 & 4 \end{pmatrix} \quad , \quad (x'x)^{-1} = C$$

إذن :

$$C = \begin{pmatrix} 1 & -1.5 \\ -1.5 & 2.5 \end{pmatrix}$$

$$\hat{\sigma}^2 = \frac{\Sigma e^2}{n-k-1}$$

وان ، $c_{11} = 1$

$$\hat{\sigma}^2 = \frac{RSS}{5-3} = \frac{1.5}{2} = 0.75$$

$$s.e(\hat{\beta}_1) = \sqrt{0.75} \cdot 1$$

إذن:

$$t^*_{\hat{\beta}_1} = \frac{2.5}{\sqrt{0.75}} = 2.89$$

– تقارن $t^*_{\hat{\beta}_1}$ مع القيمة الجدولية بمستوى دلالة 5% هو: $t_{c(2,0.025)} = 4.3036$

– القرار: إن المعلمة β_1 غير معنوية.

ولحساب فترة ثقة باحتمال 95% للمعلمة β_1 نعوض في العلاقة:

$$\hat{\beta}_1 - t_{c(\frac{\alpha}{2})} \cdot \ s.e(\hat{\beta}_1) \ \leq \ \beta_1 \ \leq \ \hat{\beta}_1 + t_{c(\frac{\alpha}{2})} \cdot \ s.e(\hat{\beta}_1)$$

$$2.5 - 4.303 \cdot \ \sqrt{0.75} \ \leq \ \beta_1 \ \leq \ 2.5 + 4.303 \cdot \ \sqrt{0.75}$$

$$2.5 - 3.727 \ \leq \ \beta_1 \ \leq \ 2.5 + 3.727$$

$$-1.23 \leq \ \beta_1 \ \leq \ 6.23$$

ويتضح أن الصفر يقع ضمن فترة الثقة للمعلمة β_1، بمعنى ان المعلمة β_1 لاتختلف معنوياً عن الصفر .

وهو النتيجة نفسها التي تم التوصل إليها عند استخدام اختبار المعنوية بالاعتماد على النسبة t.

اختبار معنوية المعلمة β_2 .

– الفرضية: $H_0: \beta_2 = 0$ vs. $H_1: \beta_2 \neq 0$

– نحسب النسبة $t^*_{\hat{\beta}_2}$.

– الانحراف المعياري للمعلمة : $\sqrt{0.75}(\sqrt{2.5}) = \hat{\sigma}\sqrt{c_{22}} = s.e(\hat{\beta}_2)$

إذن:

$$\left|t^*_{\hat{\beta}_2}\right| = \left|\frac{-1.5}{\sqrt{0.75(2.5)}}\right| = \frac{1.5}{1.369} \cong 1.1$$

– القرار: $\left|t^*_{\hat{\beta}_2}\right| \ \langle \ t_{c(0.025)}$

إذن لانرفض فرضية العدم .أي أن المعلمة β_2 لاتختلف معنوياً عن الصفر.

أما فترة الثقة باحتمال 95% للمعلمة β_2 تحسب على وفق الآتي:

$$\hat{\beta}_2 - t_{c(\frac{\alpha}{2})} \cdot s.e(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{c(\frac{\alpha}{2})} \cdot s.e(\hat{\beta}_2)$$

$$-1.5 - 4.303 \cdot (1.369) \le \beta_2 \le -1.5 + 4.303 \cdot (1.369)$$

$$-7.391 \le \beta_2 \le 4.39$$

وحيث إن الصفر يقع ضمن الفترة المذكورة (4.39 , 7.391−)، وهذا بدوره يدلل على ان المعلمة β_2لاتختلف معنوياً عن الصفر وهي منسجمة مع نتيجة الاختبار الذي اعتمد على النسبة tللمعلمة المقدرة β_2.

ولا يختلف اختبار موجبيه أو سالبيه معلمة الانحدار الجزئي في الخطوات سوى بتحديد فترة القبول أو الرفض وكما تم توضيحه في الفصل الثالث ونقطة الاختلاف هنا في تحديد درجات الحرية فهي (n−k−1) وعوضاً عن (n−2) في الانحدار البسيط.

(2−5)اختبار معنوية تركيب خطي بدلالة المعلمات. Testing significance of linear combination of parameters

قد تكون الفرضية المطلوب اختبارها هي بالصيغ التالية:

$H_0: \beta_1 = \beta_2$

$H_0: \beta_1 + \beta_2 = 1$

$H_0: \beta_1 + \beta_2 + \beta_3 = 1$ أو

$H_0: 2\beta_1 + 3\beta_2 - 5\beta_3 = 0$ أو

وبشكل عام، إذا كانت فرضية العدم:

$$H_0 : \sum_{j=1}^{k} a_j \beta_j = c$$

a_j، c ثوابت

فان مثل هكذا فرضية تسمى تركيب خطي بدلالة المعلمات .ويمكن اختبارها على وفق طريقة اختبار معلمة الانحدار وذلك باستخدام المختبر الإحصائي (نسبة t للتركيب)إذ يتم افتراض

$$\sum_{j=1}^{k} a_j \beta_j = \gamma$$

وبالتالي فالاحصاءة t تحسب على وفق الآتي:

$$t^*_{\hat{\gamma}} = \frac{\hat{\gamma}}{s.e(\hat{\gamma})}$$

ويمكن حساب الانحراف المعياري للتركيب:

$$s.e(\hat{\gamma}) = \sqrt{\text{var}(\hat{\gamma})} = \sqrt{\sum_{j=1}^{k} a_j^2 \text{ var}(\hat{\beta}_j) + 2a_i a_j \text{ cov}(\hat{\beta}_i \hat{\beta}_j)}$$

مثال :(2-5) إذا علمت أن مصفوفة العزم الثاني للمتغيرات Y و X_1 و X_2 و X_3 كانحرافات عن متوسطاتها لـ (24)مشاهدة:

$$\left[x'x \;\vdots\; x'y\right] = \left[\begin{array}{ccc:c} 10 & 10 & 5 & 7 \\ & 30 & 15 & -7 \\ & & 20 & -26 \end{array}\right]$$

م/ اختبر الفرضية: $\beta_1 + \beta_2 + \beta_3 = 0$

الحل:

$H_0: \beta_1 + \beta_2 + \beta_3 = 0$

vs.: $H_0: \beta_1 + \beta_2 + \beta_3 \neq 0$

نفرض $\beta_3 + \beta_2 + \beta_1 = \gamma$

- نسبة t للتركيب هي:

$$t^*_{\hat{\gamma}} = \frac{\hat{\gamma}}{s.e(\hat{\gamma})}$$

- $\hat{\beta}_3 + \hat{\beta}_2 + \hat{\beta}_1 = \hat{\gamma}$

- $$\text{var}(\hat{\gamma}) = \text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2\text{cov}(\hat{\beta}_1\hat{\beta}_2) + 2\text{cov}(\hat{\beta}_1\hat{\beta}_3) + 2\text{cov}(\hat{\beta}_2\hat{\beta}_3)$$

$$= \hat{\sigma}^2(c_{11} + c_{22} + c_{33} + 2c_{12} + 2c_{13} + 2c_{33})$$

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-4}}$$

يتطلب ذلك حساب المقدرات: $\hat{\beta} = (x'x)^{-1} x'y$

$$\begin{bmatrix} 10 & 10 & 5 \\ & 30 & 15 \\ & & 20 \end{bmatrix}^{-1} \begin{pmatrix} 7 \\ -7 \\ -26 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}$$

لحساب معكوس $(x'x)$ ، نحسب المحدد باستخدام طريقة سايروس.

$$|x'x| = \begin{vmatrix} 10 & 10 & 5 \\ 10 & 30 & 15 \\ 5 & 15 & 20 \end{vmatrix} = 2500$$

ثم نحسب مصفوفة المتممات cofactor :

$$cofactor = \begin{bmatrix} 375 & -125 & 0 \\ -125 & 175 & -100 \\ 0 & -100 & 200 \end{bmatrix}$$

وندورها لحساب ad joint ثم نقسم عناصر المرافقة على المحدد فنحصل على معكوس المصفوفة.

$$(x'x)^{-1} = C = \begin{bmatrix} 0.15 & -0.05 & 0 \\ -0.05 & 0.07 & -0.04 \\ 0 & -0.04 & 0.08 \end{bmatrix}$$

مقدرات النموذج بموجب OLS هي:

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{pmatrix}=(x'x)^{-1}x'y=\begin{pmatrix}1.4\\ 0.2\\ -1.8\end{pmatrix}$$

$$\Rightarrow\ \hat{\gamma}=-0.2$$

$$ESS=\hat{\beta}'x'y=\begin{bmatrix}1.4 & 0.2 & -1.8\end{bmatrix}\begin{bmatrix}7\\ -7\\ -26\end{bmatrix}=55.2$$

$$TSS=\Sigma y^2=60$$

$$RSS=TSS-ESS=4.8$$

$$\hat{\sigma}^2=\frac{RSS}{n-k-1}=\frac{4.8}{24-4}=0.24\qquad\Rightarrow\ \hat{\sigma}=0.49$$

$$\operatorname{var}(\hat{\beta}_i)=\hat{\sigma}^2c_{ii}$$

$$\Rightarrow s.e(\hat{\gamma})=0.49\sqrt{0.15+0.07+0.08-2(0.05)+0-2(0.04)}=0.17$$

نسبة t المحسوبة:

$$\left|\ t^*_{\hat{\gamma}}\right|=\left|\frac{-0.2}{0.17}\right|=1.18$$

$$t_{c(20,0.025)}=2.086$$

ومن خلال مقارنة القيمة المحسوبة t^* مع القيمة المجدولة t_c.

$$\left|\ t^*_{\hat{\gamma}}\right|\ \langle\ t_c$$

القرار: لا نرفض فرضية العدم، أي أن مجموع المعلمات الحقيقية لا تختلف معنوياً عن الصفر.

وبعبارة أخرى فان مجموع المعلمات الحقيقية للمجتمع لا تختلف معنوياً عن الصفر.

(3-5) اختبار معنوية نموذج الانحدار ككل.Testing the overall significance

أن الفرضية التي نسعى لاختبارها هي: $H_0: \beta_1 = \beta_2 = \beta_3 = \ . \ . \ . \ = \beta_k = 0$

بمعنى أن جميع الآثار المباشرة غير مهمة.

وهذه الفرضية هي فرضية مشتركة تختبر ان β_1، β_2، . . . . ، β_k تساوي صفراً بشكل آني والاختبار المستخدم هنا يسمى اختبار المعنوية الكلية للخط المستقيم المقدر. وبعبارة أخرى يتم اختبار فيما إذا كان المتغير Y مرتبط خطياً لكل من X_1 ، X_2 ، . . . . ، X_k

وفي هذا المجال لا يمكن استخدام الاحصاءة t لاختبار الفرضية المشتركة في حين تستخدم الاحصاءة F إذ أن فرضية العدم عبارة عن عدة فرضيات تتحقق آنياً. وكل من هذه الفرضيات تتأثر بالمعلومات من الفرضيات الكلية في حين اختبار t يفترض الاستقلالية بين العينات المسحوبة لاختبار كل فرضية.

وبعبارة أخرى فان اختبار سلسلة من الفرضيات المنفردة لا تتماثل مع اختبار الفرضيات نفسها بشكل آني، والسبب وراء ذلك هو أن اختبار عدة فرضيات بشكل آني تكون الفرضيات المنفردة متأثرة بالمعلومات في الفرضيات الأخرى.

ان فرضية العدم التي تتكون من عدة فرضيات مشتركة يمكن اختبارها باستخدام تحليل التباين (ANOVA).

كما تم التوضيح بان مجموع مربعات الانحرافات الكلية (TSS) تمت تجزئتها إلى جزأين وهي مجموع مربعات الانحرافات المشروحة من قبل المتغيرات التوضيحية أو مجموع مربعات الانحدار Explained)) Sum of squares(ESS)) والجزء الثاني هو مجموع مربعات الانحرافات غير المشروحة أي مجموع مربعات الخطأ أو (مجموع مربعات البواقي) Residual sum of squares(RSS)).

ولحساب جدول تحليل التباين للنموذج الخطي العام نتبع الآتي:

١- نحسب مجموع مربعات الانحرافات الكلية عن متوسطاتها:

$$TSS = S_{YY}$$

$$= Y'Y - \frac{(\Sigma Y)^2}{n} = Y'Y - n\bar{Y}^2 \qquad . \ . \ . \qquad (2-5)$$

2- نحسب مجموع المربعات العائدة للانحدار ESS باستخدام أحد الصيغ التالية:

$$\left.\begin{aligned} ESS &= \hat{Y}'\hat{Y} - n\bar{Y}^2 \\ &= \hat{\beta}'X'Y - \frac{(\Sigma Y)^2}{n} \\ &= \hat{\beta}'X'X\hat{\beta} - \frac{(\Sigma Y)^2}{n} \\ &= \hat{\beta}'C^{-1}\hat{\beta} - \frac{(\Sigma Y)^2}{n} \end{aligned}\right\} \qquad . \ . \ . \qquad (3-5)$$

٣–نحسب مجموع المربعات العائدة للخطأ (RSS):

$$e'e = Y'Y - \hat{Y}'\hat{Y}$$

$$= TSS - ESS \quad . . . \quad (4-5)$$

ثم ينظم الجدول كالآتي:

جدول(5–1)

جدول تحليل التباين

– ANOVA–

S.O.V	SS	d.f	MSS	F*
الانحدار	ESS	k	ESS / k	$\frac{EMS}{RMS}$
الخطأ	RSS	n–k–1	RSS /(n–k–1)	
الإجمالي	TSS	n–1		

أن جدول تحليل التباين في الانحدار الخطي العام يستخدم لاختبار معنوية العلاقة الخطية المستخدمة من خلال اختبار فرضية العدم:

$$H_0 : \beta_1 = \beta_2 = . . . = \beta_k = 0$$

أي أن جميع المتغيرات التوضيحية X_1 ، X_2 ، . . . . ، X_k غير مهمة في تحديد التغير في Y.

ضد الفرضية البديلة:

$$H_0 : \beta_1 \neq \beta_2 \neq . . . \neq \beta_k \neq 0$$

والتي تعرض بان على الأقل أحد المتغيرات التوضيحية (X_1 ، X_2 ، . . . . ، X_k) مهمة في تحديد التغيرات في Y. بمعنى آخر ان العلاقة الخطية ملائمة لتمثيل البيانات:

ويكون القرار بقبول أو رفض فرضية العدم بعد مقارنة القيمة العملية (المحسوبة) للاحصاءة F مع القيمة النظرية (الجدولية) للاحصاءة F على وفق مستوى دلالة مختار ودرجات حرية البسط k ودرجات حرية المقام (n–k–1).

فإذا كانت F(المحسوبة) > (F المجدولة) ولمستوى معنوية α فيكون القرار برفض H_0 . أي ان العلاقة الخطية ملائمة والعكس صحيح.

كما ويستخدم جدول تحليل التباين لمعرفة الأهمية النسبية للتغيرات المشروحة والتغيرات غير المشروحة وبالتالي يمكن حسابه بالاعتماد على معامل التحديد R^2 وعلى وفق الآتي:

جدول تحليل التباين باستخدام معامل التحديد. ANOVA Table using coefficient of determination

وحيث أن:
$$R^2 = \frac{ESS}{TSS}$$

إذن يمكن التعويض عن ($ESS = R^2\ TSS$) وكذلك

$RSS = (1-R^2)\ TSS$

في جدول تحليل التباين وبذلك فان الاحصاءة F:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \quad \ldots \quad (5\text{-}5)$$

مثال(5-3): اعتماداً على بيانات الجدول (4-3) ص١١٤، ولاختبار معنوية النموذج ككل:

$$\mathrm{H}_0 : \beta_1 = \beta_2 = 0 \quad vs. \quad \mathrm{H}_1 : \beta_1 \neq \beta_2 \neq 0$$

$$F^* = \frac{EMS}{RMS} = \frac{26.5/2}{1.5/(5-3)} = \frac{13.25}{0.75} = 17.7$$

أو باستخدام معامل التحديد:

$$F^* = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0.9464/2}{0.054/2} = \frac{0.4732}{0.0268} = 17.7$$

وتقارن مع $F_{c\,(2,2,0.95)} = 19$ وحيث أن $F^* < F_c$ ، القرار: لا نرفض فرضية العدم. أي أن العلاقة الخطية غير ملائمة للبيانات. أو أن المتغيرات X_1 و X_2 غير مهمة معنوياً.

مثال (5–4): من بيانات المثال (5–2) لاختبار معنوية النموذج ككل:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad vs. \quad H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq 0$$

جدول (ANOVA)

S.O.V	d.f	SS	MSS	F^*	$F_{c(3,20,0.95)}$
Regression	3	ESS =55.2	18.4	76.7	3.10
Error	20	RSS = 4.8	0.24		
Total	23	TSS = 60			

وبمقارنة القيمة الحسابية لـ F مع القيمة الجدولية يتضح أن $F^* > F_c$.

القرار: نرفض H_0 أي أن على الأقل أحد المتغيرات التوضيحية مهم معنوياً. أو أن العلاقة الخطية ملائمة للبيانات.

وباستخدام معامل التحديد:

$$R^2 = \frac{ESS}{TSS} = 0.92$$

يمكن إعادة جدول تحليل التباين على وفق الآتي:

جدول (ANOVA)

S.O.V	d.f	SS	MSS	F^*	$F_{c(3,20,0.95)}$
Regression	3	(0.92)(60)	18.4		
Error	20	(0.08)(60)	0.24	76.7	3.10

(5–4)مجموع المربعات الإضافي. (Additional sum of squares)

افترض النموذج الآتي:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e$$

ونرغب حساب أثر مجموعة جزئية من المتغيرات المستقلة وهي X_3 و X_4 و X_5 على التنبؤ بقيم Y. لذا تتم تجزئة المصفوفة X إلى جزأين هما X_1 و X_2.

حيث ان X_1 تشتمل على الأعمدة الأول والثاني من المصفوفة X و

X_2 تشتمل على الأعمدة الثالث والرابع والخامس من المصفوفة X.

وكذلك يتم تجزئة المتجه β إلى متجهين جزئيين يضم β_1مع β_2 والثاني يضم β_3 و β_4 و β_5 وكالاتي:

$$Y = X\beta + e$$

حيث ان:

$$\beta' = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \vdots \quad \beta_3 \quad \beta_4 \quad \beta_5]$$

$$Y = X_1\gamma_1 + X_2\gamma_2 + e$$

$$\gamma_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad , \quad \gamma_2 = \begin{bmatrix} \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}$$

ولإيجاد مقدار مساهمة المتغيرات المستقلة X_3 و X_4 و X_5 للتنبؤ لـ Y يكون من خلال أثر المتغيرات في المصفوفة الجزئية X_2 علماً ان (المتغيرات التوضيحية في المصفوفة الجزئية X_1 مضمنة أصلاً في النموذج.ويرمز لها $ESS(X_3, X_4, X_5/X_0X_1X_2)$ ويسمى مجموع المربعات الإضافي العائد إلى الحدود الموجودة في(γ_2) . وهو بذلك يقيس مجموع المربعات العائدة من إضافة المتغيرات المستقلة X_3 و X_4 و X_5 إلى النموذج والذي يحتوي على المتغيرات المستقلة X_1 و X_2 و X_0.

وتحسب $ESS(X_3X_4X_5/X_0X_1X_2)$ كالآتي:

$ESS(X_0X_1X_2X_3X_4X_5)$ للنموذج الكلي مطروحاً منه $ESS(X_0X_1X_2)$

حيث ان:

$$ESS(X_0X_1X_2X_3X_4X_5) = \beta'X'Y \quad \text{و} \quad ESS(X_0X_1X_2) = \hat{\gamma}_1X_1'Y$$

وبذلك:

$$ESS(X_3X_4X_5/X_0X_1X_2) = ESS(X_0X_1X_2X_3X_4X_5) - ESS(X_0X_1X_2)$$

وتسمى هذه الطريقة بالطريقة المباشرة لحساب مجموع المربعات الإضافية.

ويمكن حسابها بطريقة أخرى كالآتي:

$$ESS(X_3X_4X_5 \big/ X_0X_1X_2) = \begin{pmatrix} \hat{\beta}_3 & \hat{\beta}_4 & \hat{\beta}_5 \end{pmatrix} \begin{pmatrix} c_{33} & c_{34} & c_{35} \\ c_{43} & c_{44} & c_{45} \\ c_{53} & c_{54} & c_{55} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{pmatrix}$$

حيث ان:

$\begin{pmatrix} c_{33} & c_{34} & c_{35} \\ c_{43} & c_{44} & c_{45} \\ c_{53} & c_{54} & c_{55} \end{pmatrix}$ هي جزء من مصفوفة معكوس فيشر والمرتبطة بالمتغيرات X_3 ، X_4 ، X_5.

وهكذا وبشكل عام يمكن حساب مجموع المربعات المشروحة الإضافية لأي مجموعة جزئية X_i ، X_j ، X_k إلى نموذج عام بالطريقة المباشرة.

$$ESS(X_i X_j X_k / \text{جميع المتغيرات التوضيحية الأخرى}) = ESS(X_1, \ldots, X_k) - ESS(X_i X_j X_k \ \text{جميع المتغيرات عدا})$$

$$ESS(X_1/X_2) = ESS(X_1X_2) - ESS(X_2) \qquad \ldots \qquad (6\text{-}5)$$

وبالطريقة المختصرة

$$ESS(X_i X_j X_k / \text{جميع المتغيرات التوضيحية الاخرى}) = \begin{pmatrix} \hat{\beta}_i & \hat{\beta}_j & \hat{\beta}_k \end{pmatrix} \begin{pmatrix} c_{ii} & c_{ij} & c_{ik} \\ c_{ji} & c_{jj} & c_{jk} \\ c_{ki} & c_{kj} & c_{kk} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\beta}_i \\ \hat{\beta}_j \\ \hat{\beta}_k \end{pmatrix} \qquad \ldots \qquad (7\text{-}5)$$

وبالطريقة نفسها يمكن إيجاد مجموع مربعات الانحدار الإضافي العائد لمتغير مستقل واحد على وفق الآتي:

$$ESS(X_1/X_2X_3X_4X_5) = ESS(X_1X_2X_3X_4X_5) - ESS(X_2X_3X_4X_5)$$

$$= \frac{\hat{\beta}_1^{\,2}}{c_{11}}$$

$$ESS(X_2/X_1X_3X_4X_5) = \frac{\hat{\beta}_2^{\,2}}{c_{22}}$$

وبشكل عام :

$$ESS(X_i/X_1X_2, \ldots, X_{i-1}, X_{i+1}, \ldots X_K) = \frac{\hat{\beta}_i^{\,2}}{c_{ii}} \qquad \ldots \qquad (8\text{-}5)$$

مثال (5-5): مع توافر المعلومات التالية:

$$\hat{y}_i = 0.75\ x_{1i} + 3\ x_{2i} + 1.5\ x_{3i} \qquad i = 1, \ldots 15 \qquad ; \ \Sigma Y^2 = 436$$

$$\Sigma Y = 75$$

$$\left[\ x'x\ \vdots\ x'y\ \right] = \begin{pmatrix} 5.6 & -0.8 & 0.8 & \vdots & 3 \\ & 8.4 & -2.4 & \vdots & 21 \\ & & 2.4 & \vdots & -3 \end{pmatrix}$$

احسب التغيرات المشروحة من قبل المتغيرات X_1 و X_2 علماً بوجود المتغير X_3 .

الحل:

يمكن إيجاد مجموع المربعات المشروحة باستخدام الطريقة المباشرة كالآتي:

$$ESS(X_1 X_2 / X_3) = ESS(X_1 X_2 X_3) - ESS(X_3)$$

$$ESS(X_1 X_2 X_3) = \hat{\beta}' x' y$$

$$= \begin{bmatrix} 0.75 & 3 & 1.5 \end{bmatrix} \begin{pmatrix} 3 \\ 21 \\ -3 \end{pmatrix} = 60.75$$

$$ESS(X_3) = \hat{\beta}_3 \Sigma x_3 y$$

$$ESS(X_3) = \frac{S_{x_3 y}}{S_{x_3 x_3}} \cdot S_{x_3 y} = \frac{-3}{2.4}(-3) = 3.75$$

$$\Rightarrow ESS(X_1 X_2 / X_3) = 60.75 - 3.75 = 57$$

مثال(6−5): مع توافر المعلومات التالية لانحدار Y على X_2 و X_3:

$$\hat{Y} = -0.6 + X_2 + 3\ X_3 \qquad , \qquad \sum_{i=1}^{20} e_i^2 = 3.67$$

$$(X'X)^{-1} = \begin{pmatrix} 9.1 & -0.8 & -2.3 \\ & 0.16 & 0.2 \\ & & 0.7 \end{pmatrix}$$

احسب مجموع المربعات الإضافية المشروحة للمتغير X_2 علماً بتثبيت X_3 ، وللمتغير X_3 علماً بتثبيت X_2.

الحل: باستخدام الطريقة المختصرة:

$$ESS(X_2/X_3) = \frac{\hat{\beta}_2^2}{c_{22}} = \frac{1}{0.16} = 6.25$$

$$ESS(X_3/X_2) = \frac{\hat{\beta}_2^2}{c_{33}} = \frac{3^2}{0.7} = 4.2$$

(5−5)اختبار الأهمية الإضافية لمتغير معين او مجموعة جزئية من المتغيرات. Test the significancy of one variable or partial of group of additional variables

نفترض إن انحدار Y على X_2 تم تقديرها أولاً، وتم التأكد من معنوية المعلمة β_2 . ثم أضيف المتغير X_3 للنموذج، لدراسة فيما إذا كان المتغير X_3 قد أسهم في توضيح Y، والمساهمة التي يقدمها X_3 تفيد بزيادة التغيرات المشروحة لـ (Y) عند إضافة X_3 للنموذج،وكذلك زيادة R^2 بشكل معنوي نسبة إلى التغيرات غير المشروحة (RSS) . أن هذه المساهمة تسمى المساهمة الإضافية آو الحدية Marginal contribution. ومن خلال ذلك يمكن التأكد فيما إذا كان في إضافة متغير جديد للنموذج (علماً بوجود متغيرات أخرى في النموذج) مهماً. حيث أن الباحث لا يرغب بإضافة متغيرات تكون مساهمتها ضئيلة في زيادة التغيرات المشروحة.

والسؤال المطروح هنا، كيف نقرر فيما إذا كان متغير معني يسهم في تقليص RSS أي زيادة (ESS) والجواب على هذا السؤال يكمن في توسيع جدول (ANOVA) بما ينسجم مع فكرة الأهمية الإضافية.

جدول تحليل التباين لتوضيح فكرة الأهمية الإضافية لمتغير إضافي (أو متغيرات إضافية)

جدول (2-5)

S.O.V	SS	d.f	MSS
للمتغير X_2	$ESS(X_2)=\hat{\beta}_{12}^2\Sigma x_2^2$	1	$(X_2)ESS/$ ١
للمتغير X_3 مع وجود X_2	$ESS(X_3/X_2) = ESS(X_2,X_3)-ESS(X_2)$	1	$(X_3ESS/X_2)/$ ١
للمتغير X_2 و X_3 معاً	$ESS(X_2,X_3)=\hat{\beta}_2\Sigma x_2 y+\hat{\beta}_3\Sigma x_3 y$	2	$ESS\ (X_2, X_3)$ /٢
للبواقي	$RSS(X_2,X_3)$	n-3	RSS/ (n-3)
الإجمالية	$TSS=\sum y^2$	n-1	

الجدول (٥-٢) يعرض في السطر الأول معلومات عن معادلة التقدير لانحدار Y على X_2 :

$\hat{Y}_i=\hat{\beta}_1+\hat{\beta}_{12}X_{2i}$ حيث أن $\hat{\beta}_{12}$ هي القيمة المقدرة لانحدار Y على X_2 فقط، وتحسب في وفق العلاقة: $\left(\hat{\beta}_{12}=\frac{\Sigma x_2 y}{\Sigma x_2^2}\right)$

وبعد إضافة المتغير X_3 إلى النموذج الذي يتضمن X_2 وكما في السطر الثاني من الجدول فان معادلة التقدير: $\hat{Y}_i=\hat{\beta}_1+\hat{\beta}_2X_2+\hat{\beta}_3X_3$

والمعلمات $\hat{\beta}_2$ و $\hat{\beta}_3$ هي الأثر المباشر لـ X_2 و X_3 على Y وتحسب على وفق:

$$\begin{pmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{pmatrix}=(x'x)^{-1}x'y$$

أو

$$\hat{\beta}_2=\frac{\Sigma yx_2\Sigma x_3^2-\Sigma yx_3\Sigma x_2x_3}{(\Sigma x_2^2)(\Sigma x_3^2)-(\Sigma x_2x_3)^2}$$

$$\&\ \hat{\beta}_3=\frac{(\Sigma yx_3)(\Sigma x_2^2)-(\Sigma yx_2)(\Sigma x_2x_3)}{(\Sigma x_2^2)(\Sigma x_3^2)-(\Sigma x_2x_3)^2}$$

كما إن مجموع المربعات المشروحة من قبل المتغيرين X_2 و X_3 على Y تحسب على وفق:

$$ESS(X_2,X_3)=\hat{\beta}_2\Sigma x_2 y+\hat{\beta}_3\Sigma x_3 y=\hat{\beta}'x'y$$

لذلك فان التغيرات التي يتم توضيحها من قبل المتغير X_3 لانحدار $(Y=\beta_0+\beta_{12}X_2+u)$

يمكن تقديره بالفرق بين (مجموع المربعات المشروحة من قبل المتغيرين X_2 و X_3 معاً) وبين مجموع المربعات التي يتم توضيحها من قبل المتغير X_2 فقط:

$$ESS(X_3/X_2) = ESS(X_2,X_3) - ESS(X_2)$$

وبدرجات حرية مقدارها عدد المتغيرات التوضيحية المضافة للمعادلة الأصلية (وهو المتغير X_3)
أي إن متوسط المربعات المشروحة الإضافية هي:

$$\frac{ESS(X_3/X_2)}{1}$$

أما مجموع مربعات بواقي العلاقة الجديدة $RSS(X_2,X_3)$ فتحسب على وفق القانون:

$$RSS(X_2X_3) = TSS - ESS(X_2X_3)$$

وبدرجات حرية تمثل عدد المعلمات في النموذج الجديد وهي: $(n-3) \equiv n - (2+1)$

ويذلك فان قيمة F لاختبار الأهمية الإضافية هي:

$$F^* = \frac{ESS(X_3/X_2)/1}{RSS(X_2X_3)/(n-3)}$$

وتقارن مع F الجدولية بدرجات حرية (1) للبسط ، (n−3) للمقام ولمستوى دلالة α ($F_{c(1,n-3,1-\alpha)}$)
فيكون القرار بان إضافة المتغير X_3 إلى النموذج الذي يحتوي على المتغير المستقل X_2 قد ساعد على تحسين القدرة التفسيرية للنموذج إذا كانت: ($F^* > F_c$)

ويمكن تعميم ذلك لاختبار أهمية مجموعة جزئية من المتغيرات كآلاتي:

افترض إن الرغبة في حساب أثر مجموعة جزئية من المتغيرات التوضيحية وهي (X_4 و X_5 مثلاً) على التنبؤ بقيم Y. من العلاقة:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$$

فيتم تقسيم الانحدار إلى: العلاقة القديمة:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_1$$

في حين العلاقة الجديدة: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$ هي النموذج الكلي. ولحساب أهمية إضافة المتغيرين X_4 و X_5 على النموذج القديم يتبع الآتي:

النموذج الجديد: $Y = X\beta + u$

حيث ان: $\beta' = (\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5)$

يتم تقسيم معاملات الانحدار β_i إلى

$$\beta' = (\beta_1 \quad \beta_2 \quad \beta_3 \quad \vdots \quad \beta_4 \quad \beta_5) = (\gamma_1 \quad \vdots \quad \gamma_2)$$

حيث ان: $\gamma_1 = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ و $\gamma_2 = \begin{pmatrix} \beta_4 \\ \beta_5 \end{pmatrix}$

وبذلك فان معادلة الانحدار: $Y = X_1\gamma_1 + X_2\gamma_2 + u$ ، حيث X_1 يشمل الأعمدة التي تمثل المتغيرات الأول والثاني والثالث، كما أن X_2 يمثل العمودين اللذين يمثلان المتغيرات X_4 و X_5 .

ولاختبار أهمية إضافة المجموعة الجزئية γ_2 يتم اختبار الفرضية:

H_0: $\gamma_2 = 0$ H_1: $\gamma_2 \neq 0$ vs. والتي تتطلب–

حساب مجموع المربعات المشروحة الإضافية عند إضافة المتغيرات X_4 و X_5 للنموذج القديم: $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_1$. والتي يرمز لها : $ESS(X_4 X_5 / X_1 X_2 X_3)$

– والتي يتم حسابها: $ESS(\gamma_2/\gamma_1) = ESS(\gamma_1\gamma_2) - ESS(\gamma_1)$ وبدرجات حرية عددها (2) والتي تمثل عدد المتغيرات التوضيحية المضافة.

– ثم تحسب مجموع المربعات لبواقي الانحدار الجديد $RSS(X_1 X_2 X_3 X_4 X_5)$ وبدرجات حرية مقدارها: (عدد المعلمات في النموذج الجديد $(n - 6) \equiv (n -$ بضمتها المقطع الصادي β_0.

وبذلك فان المختبر المستخدم هو:

$$F = \frac{ESS(X_4 X_5 / X_1 X_2 X_3)/2}{RSS(X_1 X_2 X_3 X_4 X_5)/(n-6)}$$

أما في حال استخدام الطريقة الثانية (الطريقة المختصرة) لحساب مجموع المربعات الإضافي لحدود γ_2 فيتم على وفق الآتي:

$$ESS(X_4 X_5 / X_1 X_2 X_3) = \begin{bmatrix} \hat{\beta}_4 & \hat{\beta}_5 \end{bmatrix} \begin{bmatrix} c_{44} & c_{45} \\ c_{54} & c_{55} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix}$$

حيث $\hat{\beta}_4$ و $\hat{\beta}_5$ هي المقدرات التي يتم الحصول عليها من تقدير النموذج الكلي (الجديد)، وكذلك فان c_{44} ، c_{45} ، c_{55} هي عناصر معكوس مصفوفة فيشر للنموذج الكلي (الجديد)

أما تعميم ذلك فيتبع الآتي:

$$ESS(Z_2/Z_1) = [\hat{\gamma}_2]' \, C_{Z_2}^{-1} [\hat{\gamma}_2] \qquad . . . \qquad (9-5)$$

حيث c_{z_2} تمثل عناصر مصفوفة معكوس فيشر الخاصة بالمتغيرات Z_2.

وهكذا يمكن اختبار الأهمية الإضافية لأي مجموعة جزئية γ_2 وبأي عدد من المتغيرات.

$$F^*_{\hat{\gamma}_2/\hat{\gamma}_1} = \frac{ESS(Z_2/Z_1)/r}{RSS(Z_1,Z_2)/(n-k-1)} \quad . . . \quad (10\text{-}5)$$

حيث (r) تمثل عدد المتغيرات الموجودة في المصفوفة Z_2 (المضافة للنموذج القديم).

(k) تمثل عدد المتغيرات في النموذج الكلي (الموجودة في النموذج الجديد).

كما يمكن التأكيد على استخدام النسبة F باعتماد معامل التحديد فقط ، وذلك باستخدام العلاقة التالية:

$$F = \frac{(R^2_{new} - R^2_{old})/r}{(1-R^2_{new})/(n-k-1)} \quad . . . \quad (11\text{-}5)$$

حيث أن:

R^2_{new} : تمثل معامل التحديد للنموذج الجديد (الكلي) الذي يتضمن المتغيرات Z_1 و Z_2

R^2_{old} : تمثل معامل التحديد للنموذج القديم والذي يتضمن المتغيرات Z_1 فقط.

r : تمثل عدد المتغيرات الموجودة في المصفوفة Z_2 (المضافة للنموذج القديم)

k : تمثل عدد المتغيرات في النموذج الجديد (الكلي).

(5-6) اختبار القيود الخطية Testing Linear Equality Restrictions

في النموذج الخطي العام: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + . . . + \beta_k X_k + u_t$

ولتوضيح القيود الخطية نقدم بعض الأمثلة:

أولا:

1- $\beta_1 + \beta_2 = 1$

2- $\beta_1 + \beta_2 + \beta_3 = \frac{1}{2}$

3- $\beta_4 + \beta_5 = 0 \Leftarrow \beta_4 = \beta_5$

⋮

و هكذا

يندرج ضمنها اختبار أي تركيب خطي بدلالة المعلمات.

ويتم اختبار القيود الخطية والتي تتبع الصيغ الواردة في أعلاه بالاعتماد على اختبار t وكما تم توضيحه في فقرة اختبار التركيب الخطي. أو يمكن اختبارها على وفق اختبار F وعلى وفق الآتي:

$$F^* = \frac{(RSS_r - RSS_{Ur})/m}{RSS_{Ur}/(n-k-1)} = \frac{(e'e_r - e'e_{Ur})/m}{e'e_{Ur}/(n-k-1)} \quad . . . \quad (12\text{-}5)$$

حيث إن : $RSS_r = e'e_r$: يمثل مجموع مربعات الأخطاء لعلاقة الانحدار بعد إدخال القيد المفروض(علاقة الانحدار المقيدة والتي تحقق فرضية العدم)

$RSS_{Ur} = e'e_{Ur}$: تمثل مجموع مربعات الأخطاء لعلاقة الانحدار الكلي (غير المقيد)

m :تمثل عدد القيود m=1 .

ولتوضيح الفكرة:

إذا افترضنا القيد $\beta_1 + \beta_2 = 1$

تعاد صياغته كالأتي: $\beta_1 = (1-\beta_2)$ وتعوض هذه القيمة في النموذج الأصلي، فيصبح النموذج المقيد:

$$Y = \beta_0 + (1-\beta_2)X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

$$Y = \beta_0 + X_1 - \beta_2 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

$$(Y - X_1) = \beta_0 + \beta_2(X_2 - X_1) + \beta_3 X_3 + \ldots + \beta_k X_k + u$$

نفرض: $Y - X_1 = Y^*$ و $X_2 - X_1 = X_2^*$ ، $X_3 = X_3^*$ ، ... ، $X_k = X_k^*$

وبذلك فالنموذج المقيد: $Y^* = \alpha_1 + \alpha_2 X_2^* + \alpha_3 X_3^* + \ldots + \alpha_k X_k^* + v$

ويتم حساب مجموع المربعات المشروحة للنموذجين (الكلي) والنموذج المقيد، وكذلك مجموع المربعات غير المشروحة لكلا النموذجين وتطبيق العلاقة (5-12).

وجدير بالذكر إن العلاقة (5-12) يمكن إعادة صياغتها على وفق القانون الآتي:

$$F^* = \frac{(ESS_{Ur} - ESS_r)/m}{RSS_{Ur}/(n-k-1)} \quad \ldots \quad (13\text{-}5)$$

حيث أن: $RSS = TSS - ESS$

كما يمكن استخدام معامل التحديد للنموذج الأصلي (R_{Ur}^2) والنموذج المقيد (R_r^2) لحساب النسبة F وكالاتي:

$$F^* = \frac{(R_{Ur}^2 - R_r^2)/m}{(1-R_{Ur}^2)/(n-k-1)} \quad \ldots \quad (14\text{-}5)$$

R_{Ur}^2 : تمثل معامل التحديد للنموذج غير المقيد والذي يمثل النموذج الجديد الذي يحوي جميع المتغيرات.

R_r^2 : تمثل معامل التحديد للنموذج المقيد الذي يحقق فرضية العدم والذي يتضمن المتغيرات X_1 ، X_2 ، X_3 وهو النموذج القديم.

مع ملاحظة انه لا يمكن استخدام معامل التحديد للمقارنة بين نموذجين إلا في حالة كون قيم المتغير المعتمد في النموذجين متماثلة وكذلك عدد المشاهدات.

ثانياً : في حاله كون فرضية العدم تتضمن أكثر من فرضية خطية مستقلة .

مثال (1)

$\beta_1 = \beta_2$

و $\beta_2 + \beta_3 + \beta_4 = 1$

فيمكن كتابتها :

$H_0: \beta_1 - \beta_2 = 0$

و $\beta_2 + \beta_3 + \beta_4 = 1$

وتعاد صياغتها بالصيغة المصفوفية :

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

مثال: (2) $H_0: \beta_1 = 0$

و $\beta_2 = 0$

$\beta_4 + \beta_5 = 1$

يمكن إعادة كتابتها :

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

وهكذا: $H_0 = \beta_1 = \beta_2 = \dots\dots\dots\dots = \beta_k = 0$

يمكن إعادة كتابتها :

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

وبشكل عام فان القيود الخطية تتبع الآتي: $R\beta = r$

R:مصفوفة معاملات المعلمات في القيود الخطية. ويكون ترتيبها j × k اذ‌تمثل عدد القيود الخطية المطلوب اختبارها .

K: عدد المتغيرات المستقلة في النموذج.

β: متجه معلمات النموذج بدون المقطع الصادي : $\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$ وبترتيب (k × 1)

r : متجه القيم المطلقة في معادلات القيود وبترتيب (j × 1)

ولاختيار الفرضية الخطية العامة RB = r نتبع الآتي :

1- التأكد من أن جميع القيود الخطية تكون مستقلة بعضها عن البعض. ويتم حذف القيود غير المستقلة.

2- تعوض القيود المستقلة بالمعادلة الأصلية للحصول على المعادلة المقيدة .

3- تحسب ESS لكل من المعادلة الكلية (الأصلية) والمعادلة المقيدة أو RSS لكليهما .

4- يتم تطبيق المعادلة (5 –12) أو (5 – 13) .

5- أو تحسب R^2 للمعادلة الكلية والمعادلة المقيدة ويتم تطبيق العلاقة (5 – 14) لحساب النسبة F مع التأكيدعلى أن قيم المتغير المعتمد في النموذج المقيد وغير المقيد هي نفسها وكذلك حجم العينة.

6- تقارن مع Fالجدولية بدرجات حرية البسط (عدد القيود المستقلة=m) ودرجات حرية المقام n-(k-1)

<u>أمثله توضيحية:</u>

مثال(5-7): افترض النموذج التالي كانحرافات عن المتوسط : $Y = \beta_1 x_1 + \beta_2 x_2 + u$

وتم توفير المعلومات حول العينة :

$$\Sigma x_1^2 = 30 \quad ، \quad \Sigma y^2 = \frac{493}{3} \quad ، \quad n = 100$$

$$\Sigma x_2^2 = 3 \quad ، \quad \Sigma x_1 y = 30 \quad ، \quad \Sigma x_2 y = 20 \quad \Sigma x_1 x_2 = 0$$

م 1/- قدر معلمات النموذج باستخدام OLS واحسب R^2 .

2- اختبر الفرضيات التالية:

(أ) $VSH_0: \beta_2 = 7 \qquad H_1: \beta_2 \neq 7$

(ب) $VSH_0: \beta_1 = \beta_2 = 0 \qquad H_1: \beta_1 \neq \beta_2 \neq 0$(ج)

$H_0: \beta_2 = 7\beta_1 \qquad VS \qquad H_1: \beta_2 \neq 7\beta_1$

$$(x'x) = \begin{pmatrix} 30 & 0 \\ 0 & 3 \end{pmatrix} \quad ; \quad x'y = \begin{pmatrix} 30 \\ 20 \end{pmatrix}$$

الحل (1)

$$(x'x)^{-1} = \begin{pmatrix} \frac{1}{30} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{20}{3} \end{pmatrix}$$

$$R^2 = \frac{ESS}{TSS} \quad ; \quad ESS = \hat{\beta}'x'y = \begin{pmatrix} 1 & \frac{20}{3} \end{pmatrix} \begin{pmatrix} 30 \\ 20 \end{pmatrix} = \frac{490}{3}$$

$$\therefore \quad R^2 = \frac{490/3}{493/3} = \frac{490}{3} \times \frac{3}{493} = 0.994$$

(2) (أ) $H_0: \beta_2 = 7 \qquad VSH_1: \beta_2 \neq 7$

يمثل اختبار معلمة انحدار منفردة ويتم اختبارها باستعمال الاحصاءةt:

$$t^* = \frac{\hat{\beta}_2 - 7}{s.e(\hat{\beta}_2)}$$

$$\text{var}(\hat{\beta}_2) = \frac{RSS}{n-3} \cdot c_{22} \qquad , RSS = TSS - ESS = 1$$

$$= \frac{1}{97} \cdot \frac{1}{3} = \frac{1}{291} = 0.00344 \Rightarrow s.e(\hat{\beta}_2) = 0.0587$$

$$\left| t^*_{\hat{\beta}_2} \right| = \left| \frac{\frac{20}{3} - 7}{0.0587} \right| = 5.679 \qquad , \qquad t_c = 2$$

القرار : نرفض H_0

أيأن: β_2تختلف معنويا عن (7)

(ب) $H_0 : \beta_1 = \beta_2 = 0 \quad vs \quad H_1 : \beta_1 \neq \beta_2 \neq 0$

يمثل اختبار أكثر من فرضية في آن واحد ويتم بذلك اختبارها على وفق المختبر F :

$$F^* = \frac{ESS(x_1x_2)/2}{RSS(x_1x_2)/(n-3)} = \frac{\left(\frac{490}{3}\right)/2}{1/97} = \frac{490*97}{6} \approx 7921.67$$

القرار نرفض H_0أيأن العلاقة الخطية ملائمة للبيانات المستخدمة.

ويمكن استخدام R^2 لحساب F

$$F^* = \frac{R^2/2}{(1-R^2)/(n-3)} = \frac{(490/493)/2}{(1-490/493)/97} \approx 7921.67$$

$H_0: \beta_2 = 7\beta_1 VS H_1: \beta_2 \neq 7\beta_1$(ج)

$$\beta_2 - 7\beta_1 = 0 \qquad\qquad \beta_2 - 7\beta_1 \neq 0$$

يمكن اعتبارها كتركيب خطي $\gamma = \beta_2 - 7\beta_1$وبذلك يتم اختبارها باستخدام الاحصاءة t :

$$t^* = \frac{\hat{\beta}_2 - 7\hat{\beta}_1}{s.e(\hat{\beta}_2 - 7\hat{\beta}_1)}$$

$$\text{var}(\hat{\beta}_2 - 7\hat{\beta}_1) = \text{var}(\hat{\beta}_2) + 49\,\text{var}(\hat{\beta}_1) - 14\,\text{cov}(\hat{\beta}_1, \hat{\beta}_2)$$

$$= \frac{1}{97}(c_{22} + 49c_{11} - 14c_{12})$$

$$= \frac{1}{97}(\frac{1}{3} + 49\frac{1}{30} - 14(0)) = 0.0203$$

$$s.e(\hat{\beta}_2 - 7\hat{\beta}_1) = 0.1425$$

$$|\ t^*| = \left|\frac{\frac{20}{3} - 7}{0.1425}\right| = \left|\frac{-\frac{1}{3}}{0.1425}\right| = 2.339 \quad \rangle \quad t_{c(97,0.025)}$$

القرار : نرفض فرضية العدم .

كما يمكن اختبارها باستخدام القيود الخطية $R\beta = r$:

حيث أن $R = [-7 \quad 1]$

وباستخدام العلاقة ($5 - 12$) ، حيث m =1، r=0

نعوض $\beta_2 = 7\beta_1$ في المعادلة الأصلية :

النموذج المقيد:

$$y=\beta_1 x_1+7\beta_1 x_2+u$$

$$y=\beta_1(x_1+7x_2)+u$$

$$\therefore \hat{\beta}_1=\frac{\Sigma(x_1+7x_2)y}{\Sigma(x_1+7x_2)^2}=\frac{\Sigma x_1 y+7\Sigma x_2 y}{\Sigma x_1^2+14\Sigma x_1 x_2+49\Sigma x_2^2}$$

$$=\frac{30+140}{30+0+49(3)}=\frac{170}{177}=0.96$$

$$ESS_r=\hat{\beta}_1\Sigma(x_1+7x_2)y=163.2 \quad \Rightarrow \quad e'e_r=RSS_r=\frac{493}{3}-163.2=1.13$$

$$ESS_{Ur}=163.333 \quad \Rightarrow \quad e'e_{Ur}=1$$

وبتطبيق العلاقة $(12-5)$: $F^*=\frac{(1.13-1)/1}{1/97}=\frac{0.13}{0.0103}=12.621$

القرار : نرفض H_0

مثال(5−8): إذاأعطيت معادلة التقدير $\hat{y}=0.75x_1+3x_2+1.5x_3$ ومصفوفة العزم الثاني لانحرافات Y و X_1 و X_2 وX_3 عن متوسطاتها لعينة من (15) مشاهدة:

$$[x'x \;\vdots\; x'y]=\begin{bmatrix} 5.6 & -0.8 & 0.8 & \vdots & 2 \\ & 8.4 & -2.4 & \vdots & 21 \\ & & 2.4 & \vdots & -3 \end{bmatrix}$$

$$\sum Y=75 \quad , \quad \sum Y^2=436$$

م/(1) احسب التغيرات المشروحة من قبل X_1 و X_2 مع تثبيت X_3 .

(2) احسب الأهمية النسبية للتغيرات المشروحة من قبل المتغيرين X_1 و X_2 علما بوجود المتغير X_3

(3) اختبر الأثر الكلي للمتغير X_3 .

(4) اختبر الأثر الإضافي للمتغير X_3 .

(5) اختبر الأهميةالإضافية للمتغيرين $(X_2 \;,\; X_1)$ على Yعلماً بوجود المتغير X_3 .

(6) اختبر الفرضية : $R\beta = 0$: H_0 علماَ بان :

$$R = \begin{pmatrix} 0 & 2 & 0-2 \\ 0 & 0 & 1\ 0 \\ 0 & 2 & 1-2 \\ 0 & 2 & 3-2 \end{pmatrix}$$

(7) احسب معامل التحديد للنموذج الكلي والنموذج المقيد .

(8) اختبر الفرضية : $R\beta = 0$ حيث أن

$$R = \begin{pmatrix} 0 & 0 & 0\ 0 \\ 0 & 1 & 0\ 0 \\ 0 & 2 & 2-2 \\ 0 & 3 & 2-2 \end{pmatrix}$$

الحل: (1)

$$ESS(X_1X_2/X_3) = ESS(X_1X_2X_3) - ESS(X_3)$$

$$ESS(X_1X_2X_3) = \hat{\beta}'\ \acute{x}y = [0.75 \quad 3 \quad 1.5] \begin{bmatrix} 2 \\ 21 \\ -3 \end{bmatrix}$$

$$= 1.5 + 63 - 4.5 = 60$$

$$ESS(X_3) = \frac{\Sigma x_3 y}{\Sigma x_3^2} \cdot \Sigma x_3 y = \frac{-3}{2.4} \cdot (-3) = \frac{9}{2.4} = 3.75$$

$$\therefore \quad ESS(X_1X_2/X_3) = 60 - 3.75 = 56.25$$

أيأن56.25% من التغيرات الإجمالية بالمتغير Yيتم توضيحها عند استعمال المتغيرين X_1 و X_2في النموذج واعتبار X_3ثابت .

(2) الأهمية النسبية للتغيرات المشروحة من قبل المتغيرين X_1 و X_2 علماَ بوجود المتغير X_3 يتم حسابها بالاعتماد على العلاقة.

$$r^2_{X_1X_2.X_3} = \frac{ESS(X_1X_2/X_3)}{RSS(X_3)}$$

$$RSS(X_3) = TSS - ESS(X_3)$$

$$RSS(X_3) = TSS - ESS(X_3)$$

$$TSS = \sum Y^2 - \frac{(\sum Y)^2}{n} = 436 - 375 = 61$$

$$ESS(X_3) = 3.75 \quad \rightarrow RSS(X_3) = TSS - ESS(X_3) = 61 - 3.75$$

$$= 57.25$$

$$\therefore \quad r^2_{X_1X_2.X_3} = \frac{56.25}{57.25} = 0.983$$

(٣)الأثر الكلي للمتغير X_3 يتم حسابه من الانحدار البسيط إذ انه يمثل الأثر المباشر وغير المباشر للمتغير X_3 على المتغير Y .

$$\hat{\beta}_3 = \frac{\sum x_3 y}{\sum x_3^2} = 1.25$$

$$H_0 = \beta_3 = 0 \qquad VS. \qquad H_1 = \beta_3 \neq 0$$

$$s.e(\hat{\beta}_3) = \sqrt{\frac{RSS}{n-2}} \cdot \sqrt{\frac{1}{\sum x_3^2}}$$

$$RSS(X_3) = 57.25 \quad \Rightarrow \quad \frac{RSS(X_3)}{n-2} = 4.4$$

$$s.e(\hat{\beta}_3) = \sqrt{\frac{4.4}{2.4}} = 1.35$$

$$t^*_{\hat{\beta}_3} = \frac{\hat{\beta}_3}{s.e(\hat{\beta}_3)} = \frac{1.25}{1.35} = 0.926$$

تقارن مع الجدولية : $\quad t_c\ (13, 0.025) = 2.160$

القرار :لا نرفض H_0إيأن المتغير X_3 غير مهم معنويا في تحديد التغيرات للمتغير المعتمد Y .

(4) الأثر الإضافي يعني اختبار أهمية إضافة المتغير X_3إلى النموذج القديم الذي يحتوي على X_1 مع X_2 فقط وتحسب على وفق المختبر F العلاقة (9−5) :

$$F^* = \frac{ESS(X_3/X_1X_2)/1}{RSS(X_1X_2X_3)/(n-k-1)}$$

$$RSS(X_1X_2X_3) = TSS - ESS(X_1X_2X_3) = 61 - 60 = 1$$

$$\Rightarrow \frac{RSS(X_1X_2X_3)}{n-4} = \frac{1}{11} = 0.091$$

$$ESS(X_1X_2X_3) = ESS(X_1) + ESS(X_2/X_1) + ESS(X_3/X_1X_2)$$
$$= ESS(X_1, X_2) + ESS(X_3/X_1X_2)$$

$$\Rightarrow ESS(x_1x_2) = \hat{\beta}\ x'y$$

اولاً: يجب تقدير $\hat{\beta}_1$و $\hat{\beta}_2$ من انحدار Y ضد X_1 و X_2 فقط:

$$(x'x)_{12} = \begin{pmatrix} 5.6 & -0.8 \\ & 8.4 \end{pmatrix} \Rightarrow (x'x)^{-1} = \begin{bmatrix} 0.18 & 0.017 \\ & 0.12 \end{bmatrix}$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0.717 \\ 2.554 \end{pmatrix}$$

$$ESS(X_1X_2) = (\ 0.717 \quad 2.554\ ) \begin{pmatrix} 2 \\ 21 \end{pmatrix} = 55.068$$

$$ESS(X_3 / X_1X_2) = 60 - 55.068 = 4.932$$

وبتطبيق العلاقة (5−9) نحصل على :

$$F^* = \frac{4.932}{0.091} = 54.198$$

القرار: أنالأهميةالإضافية للمتغير X_3 معنوية . أيأنإضافة المتغير X_3 للنموذج الذي يحتوي (X_2 , X_1) ستعزز القدرة التنبؤية للنموذج .

(5)لاختيار الأهميةالإضافية للمتغيرين (X_2 , X_1) على Y علماً بوجود X_3 .

أيأن النموذج القديم هو $Y= f\ (X_3)$ ، ثم تمتإضافة المتغيرين X_2 , X_1 فيصبح النموذج الجديد$Y=f(X_1X_2X_3)$

ولاختيار الأهميةالإضافية للمتغيرين يتم تطبيق العلاقة (5−9) : حيث r = 2 عدد المتغيرات المضافة للنموذج القديم. k = 3 عدد المتغيرات في النموذج الجديد .

$$ESS(X_1X_2 / X_3) = 56.25$$

$$F^* = \frac{ESS(X_1X_2/X_3)/2}{RSS(X_1X_2X_3)/(n-k-1)} = \frac{(56.25)/2}{0.091} = \frac{28.125}{0.091} = 309.07$$

القرار : الأهميةالإضافية للمتغيرين X_1 ,X_2 مهمة معنوياً .

(6) لاختبار الفرضية الخطية العامة :

$$\begin{bmatrix} 0 & 2 & 0 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 2 & 1 & -2 \\ 0 & 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

أي اختبار الفرضيات الخطية : $2\beta_1 - 2\beta_3 = 0 \quad (1) \Rightarrow \beta_1 = \beta_3$

(2) $\beta_2 = 0$

(3) $2\beta_1 + \beta_2 - 2\beta_3 = 0$

$2\beta_1 + 3\beta_2 - 2\beta_3 = 0 \quad (4)$

ويتضح أن الفرضيات (3) و (4)هي تركيب خطي بدلالة الفرضيات (1) و (2)

فالفرضية (3) هي الفرضية (1) + الفرضية (2)

والفرضية (4) هي الفرضية (1) + 3* (الفرضية 2)

وعلية فان الفرضيات المستقلة هي (1) و (2) ، تعوض في النموذج الكلي .

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

فيصبح النموذج المقيد:

$$y = \beta_3 x_1 + \beta_3 x_3 + u$$
$$y = \beta_3 (x_1 + x_3) + u$$

نستخرج ESS لكل من النموذج الأصلي والنموذج المقيد .

$$ESS_r = \hat{\beta}_3 \Sigma(x_1 + x_3) y = \frac{(\Sigma(x_1 + x_3) y)^2}{\Sigma(x_1 + x_3)^2} = \frac{(\Sigma x_1 y + x_3 y)^2}{\Sigma x_1^2 + \Sigma x_3^2 + 2\Sigma x_1 x_3}$$

$$= \frac{(2 + (-3))^2}{5.6 + 2(0.8) + 2.4} = \frac{1}{9.6} = 0.104$$

$$ESS_{Ur} = (\hat{\beta}_1 \quad \hat{\beta}_2 \quad \hat{\beta}_3) x'y = (0.75 \quad 3 \quad 1.5) \begin{pmatrix} 2 \\ 21 \\ -3 \end{pmatrix} = 1.5 + 63 - 4.5 = 60$$

$$ESS_{Ur} - ESS_r = 59.896$$

وبتطبيق العلاقة (13-5)

$$F^* = \frac{(59.896)/2}{0.091} = \frac{29.948}{0.091} = 329.099$$

حيث أن $r = 2$ وهي عدد القيود الخطية المستقلة في فرضية العدم .

ويمكن حسابها كالأتي:(عدد المعلمات في النموذج الكلي)− عدد المعلمات في النموذج المقيد

3−1=2

القرار: أن القيود المتضمنة في فرضية العدم تساعد على تحسين القدرة التنبؤية للنموذج.

(7) معامل التحديد للنموذج الكلي= $\frac{ESS(X_1X_2X_3)_{Ur}}{TSS}$

$$R^2_{Ur} = \frac{60}{61} = 0.984$$

معامل التحديد للنموذج المقيد = $\frac{ESS_r}{TSS}$

$$R^2_r = \frac{0.104}{61} = 0.0017$$

كما يمكن اختبار الفرضية الخطية العامة بالاعتماد على معاملات التحديد للنموذج الكلي والنموذج المقيد وذلك بتطبيق العلاقة (5−14)

$$F^* = \frac{(0.984-0.0017)/2}{(1-0.984)/(n-4)} = \frac{0.49115}{0.0015} = 327.4$$

والنتيجة غيرمنطبقة بسبب التقريب الذي تم استخدامه .

غيران الاستنتاج النهائي متطابق تماماً . كما في استخدام العلاقة (5-4) .

(8) لاختبار الفرضية العامة :

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 2 & -2 \\ 0 & 3 & 2 & -2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

تعني :

$$\beta_1 = 0 \qquad (1)$$

$$2\beta_1 + 2\beta_2 - 2\beta_3 = 0 \qquad (2) \qquad \Rightarrow \qquad 2\beta_2 - 2\beta_3 = 0 \Rightarrow 2\beta_2 = 2\beta_3$$

$$3\beta_1 + 2\beta_2 - 2\beta_3 = 0 \qquad (3)$$

تصبح الفرضية (3) غير مستقلة ،وبذلك فان عدد الفرضيات المستقلة هي (2).

تعوض الفرضيات المستقلة في النموذج الأصلي فنحصل على النموذج المقيد.

النموذج غير المقيد(الأصلي):

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

النموذج المقيد:

$$y = \beta_3 x_2 + \beta_3 x_3 + u$$

$$y = \beta_3 (x_2 + x_3) + u$$

نحسب مجموع المربعات المشروحة للنموذج المقيد :

$$ESS_r = \frac{[\Sigma(x_2 + x_3)y]^2}{\Sigma(x_2 + x_3)^2} = \frac{[21 + (-3)]^2}{(8.4) \;\; + (2.4) \;\; - 2(2.4)} = \frac{(18)^2}{6} = \frac{324}{6} = 54$$

$$ESS_r = 54$$

$$ESS_{ur} = 60$$

$$\frac{RSS}{n-4} = 0.091$$

وبتطبيق العلاقة (5-13):

$$F^* = \frac{(60 - 54)/2}{0.091} = \frac{3}{0.091} = 32.967$$

أيأن القيود تساعد على تحسين القدرة التنبؤية للنموذج .

مثال(5-9): تم تقدير انحدار Y على X_1 فتم الحصول على النتائج

$$\hat{Y} = 12.762 + 0.8812\, X_1$$
$$t \qquad 2.725 \qquad 77.29$$

$n = 15$ ، $r^2 = 0.9978$

TSS = 66042.2693

$ESS(X_1) = 65898.235$

ثم أضيف المتغير X_2 فأعطى النتائج :

$$\hat{Y} = 53.16 + 0.7266\, X_1 + 273\, X_2$$
$$s.e \qquad (0.0487) \qquad (0.848)$$

$$\bar{R}^2 = 0.9986 \quad R^2 = 0.9988$$

فكم من التغيرات أسهم X_2 في توضيحها من إجمالي التغيرات في y علماً بوجود X_1 ؟

الحل :

$ESS(X_1) = 65898.235$

$ESS(X_2 / X_1) = E(X_1\, X_2) - E(X_1)$

$ESS(X_1\, X_2) = TSS \,.\, R^2 = 65963.018$

$ESS(X_2 / X_1) = 64.7836$ ، $RSS(X_1\, X_2) = TSS - ESS(X_1\, X_2) = 79.2513$

ولاختبار معنوية المساهمة الحدية للمتغير X_2 :

$$F^* = \frac{ESS(X_2/X_1)/1}{RSS(X_1X_2)/(n-3)} = \frac{64.7836/1}{79.2513/12} = 9.81$$

وكما يمكن الاختبار بالاعتماد على معامل التحديد للنموذج الأصلي والنموذج المقيد .

$$F^* = \frac{(R_{new}^2 - R_{old}^2)}{(1 - R_{new}^2)/(n-3)} = \frac{0.9988 - 0.9978}{0.0001} \cong 10$$

وهي تحدد أهمية إضافة المتغير X_2 .

(7-5) مجال الثقة لمتوسط الاستجابة وللقيمة التنبؤية الجديدة.

Cofidence interval for E(Y/X$_0$) and for new prediction.

في مباحث سابقة تم توضيح مجال الثقة لمعلمة مقدرة منفردة (β_i) يمكن تحديده اعتماداً على التوزيع (t) :

$$t_{(\hat{\beta}_i)} = \frac{\hat{\beta}_i - \beta_{i0}}{s.e(\hat{\beta}_i)} \sim t_{(n-k-1)}$$

أما مجال الثقة المشترك لمعلمتين أو أكثر يمكن تحديده اعتماداًعلى معادلة القيود الخطية :

$R\beta = r$ وذلك بتحديد المصفوفة R :

وحيث أن $\hat{\beta}$ تمت برهنتها بأنها تتوزع توزيعاً طبيعياً بمتوسط (β) وتباين $\sigma^2 c$

$$\hat{\beta} \sim N(\beta, \sigma^2 c)$$

$$\Rightarrow \qquad R\hat{\beta} \sim N(R\beta, \sigma^2 R\, C\, \acute{R})$$

حيث :

$$E(R\hat{\beta}) = R\,\beta$$

$$\text{var}(R\hat{\beta}) = E\left(R(\hat{\beta}-\beta)(\hat{\beta}-\beta)'R'\right)$$

$$= \sigma^2 R\, C\, \acute{R}$$

فعند اختبار المعنوية المشتركة للمعلمين β_1 , β_2 معاً فان فرضية العدم :

$$H_0 = \beta_1 = \beta_2 = 0 \qquad \Leftrightarrow \beta_1 = 0 \ \ \& \beta_2 = 0$$

أيأن : $\qquad R\beta = r$

يمكن كتابتها كالأتي :

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ومع تحققفرضية العدم:

$$R(\hat{\beta}-\beta)=R\hat{\beta}-R\beta=R\hat{\beta}-r\sim \mathrm{N}(0,\sigma^2 RCR')$$

$$\Rightarrow (R\hat{\beta}-r)'\left[\sigma^2 RCR'\right]^{-1}(R\hat{\beta}-r)\sim \chi_q^2$$

ولكن: $\frac{(R\hat{\beta}-r)'\left[RCR'\right]^{-1}(R\hat{\beta}-r)/q}{e'e/(n-k)} \sim F_{(q,n-k)} \Leftarrow \frac{e'e}{\sigma^2}\sim \chi_{n-k}^2$

أما لتحديد مجال الثقة التنبؤي لـ Y نسبة إلىX_{10} و X_{20} خارج حدود العينة فيعتمد على القيمة التنبؤية لـ Y والتي يتم الحصول عليها بتعويض X_{10} و X_{20} في معادلة التقدير: $\hat{Y}=R\hat{\beta}$

حيث ان: $R=[1X_{10}X_{20}]$

والقيمة الحقيقية في الفترة التنبؤية ستكون :

$$Y_f = R\beta + u_f$$

u_f: تمثل القيمة الحقيقية المفترضة للمتغير العشوائي في الفترة التنبؤية .وعلية فيمكن تعريف خطأ التنبؤ(e_f)والذي يحسب على وفق الآتي:

$$e_f = Y_f - \hat{Y}_f$$
$$= -R(\hat{\beta}-\beta)+u_f$$

وواضح أن$E(e_f)=0$

حيث أن $E(\hat{\beta})=\beta$ وكذلك$E(u_f)=0$

وبذلك:

$$V(e_f)=E\left[-R(\hat{\beta}-\beta)+u_f\right]\left[-R(\hat{\beta}-\beta)+u_f\right]'$$

$$=\sigma^2 R(X'X)^{-1}R'+\sigma^2=\sigma^2\left[R(X'X)^{-1}R'+1\right]$$

$$\mathrm{var}(R\beta)=\sigma^2 R(X'X)^{-1}R'$$

لان :

$$e_f = Y_f - \hat{Y}_f \sim N\left(0, \sigma^2\left[R(X'X)^{-1}R'+1\right]\right)$$

وبذلك :

$$\frac{Y_f-\hat{Y}_f}{\sigma\sqrt{1+R(X'X)^{-1}R'}}\sim N(0,1)$$

وبالتعويض عن قيمة σ بقيمتها المقدرة :

$$s = \hat{\sigma} = \sqrt{\frac{e'e}{n-k-1}}$$

فان :

$$\frac{Y_f - \hat{Y}_f}{\hat{\sigma}\sqrt{1 + R(X'X)^{-1}R'}} \quad \sim \quad t_{(n-k-1,\frac{\alpha}{2})}$$

وبذلك فان مجال الثقة باحتمال 95% لـ Y_f يكون :

$$\hat{Y}_f \pm t_{0.025}\hat{\sigma}\sqrt{1 + R(X'X)^{-1}R'}$$

أما فيما يخص مجال الثقة لمتوسط الاستجابة $E(Y_f)$: القيمة المتوقعة لـY في الفترة التنبؤية .

$$Y_f = R\beta + u_f$$

$$E(Y_f) = R\beta$$

$$e_f = E(Y_f) - \hat{Y}_f = -R(\hat{\beta} - \beta)$$

وبالتالي فان 95%مجال ثقة لمتوسط الاستجابة :

$$\hat{Y}_f \pm t_{0.025} \cdot \hat{\sigma}\sqrt{R(X'X)^{-1}R'}$$

والتي تكون أضيق من مجال الثقة للقيمة التنبؤية الجديدة Y_f .

<u>مثال(5-10)</u>:

$$(X'X)^{-1} = \begin{bmatrix} 26.7 & 4.5 & -8.0 \\ & 1.0 & -1.5 \\ & & 2.5 \end{bmatrix} \quad , \quad \hat{Y} = 4 + 2.5X_1 - 1.5X_2 \quad , \quad n = 5$$

$$\hat{\sigma}^2 = 0.75$$

احسب مجال الثقة باحتمال 95% للقيمة التنبؤية الجديدة عندما تكون$X_1=10$ و $X_2=10$ وكذلك لمتوسطالاستجابة .

الحل :

$$R(X'X)^{-1}R' = [1 \quad 10 \quad 10]\begin{bmatrix} 26.7 & 4.5 & -8.0 \\ & 1.0 & -1.5 \\ & -1.5 & 2.5 \end{bmatrix}$$

$$= [-8.3 \quad -0.5 \quad 2.0] \begin{bmatrix} 1 \\ 10 \\ 10 \end{bmatrix} = 6.7$$

$$\hat{Y}_f = 14$$

$t_{(0.025,2)} = 4.303$

وبذلك فان مجال الثقة باحتمال %95 لمتوسط الاستجابة :

$$14 \pm 4.303\sqrt{0.75}\sqrt{6.7}$$

والمتمثل بالفترة: $(23.646 \quad , \quad 4.354)$

ويمكن استخدام البيانات كانحرافات عن متوسطاتها وعلى وفق الآتي: المتجه X_f كانحرافات:

$$x_f = \left[X_{10} - \bar{X}_1 \qquad X_{20} - \bar{X}_2\right]$$

وتكون القيمة المقدرة لتباين متوسط الاستجابة لـ Y_f :

$$\text{var}(\hat{Y}_f) = \hat{\sigma}^2\left[\frac{1}{n} + x_f(x'x)^{-1}x'_f\right]$$

اما القيمة المقدرة لتباين القيمة التنبؤية لـ Y_f :

$$\text{var}(\hat{Y}_f) = \hat{\sigma}^2\left[1 + \frac{1}{n} + x_f(x'x)^{-1}x'_f\right]$$

(8−5)معامل الانحدارالجزئي القياس: Standard Partial regressio coefficient

يرمز لهذا المعامل β_i^* ، وهو معامل الانحدار الجزئي عندما يتم تحويل متغيرات النموذج الى صيغتها القياسية كالأتي :

$$X_{ji}^* = \frac{X_{ji}-\overline{X}_j}{s.e(X_j)} \qquad \forall \qquad j = 1, \dots\dots, k$$

وكذلك :

$$i =, \dots\dots, n$$

$$Y_i^* = \frac{Y_i-\overline{Y}}{s.e(Y)} \qquad i = 1, \dots\dots\dots, n$$

حيث ان :

$$s.e(X_j) = \sqrt{\frac{\Sigma x_j^2}{n-1}} = \sqrt{\frac{\Sigma(X_{ji}-\overline{X})^2}{n-1}}$$

وكذلك :

$$s.e(Y) = \sqrt{\frac{\Sigma(Y_i-\overline{Y})^2}{n-1}}$$

وبذلك فان معادلة الانحدار القياسية تكون خالية من المقطع الصادي :

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_m^* X_{ki}^* + e_i$$

كما أن المعلمات القياسية المقدرة بطريقة المربعات الصغرى :

$$\hat{\beta}^* = (\acute{X}^* \ X^*)^{-1} \ \acute{X}^* Y^*$$

ويمكن إعادة صياغتها بدلالة المعلمات الجزئية غير القياسية :

$$\hat{\beta}_j = \hat{\beta}_j^* \cdot \sqrt{\frac{S_{yy}}{S_{x_j x_j}}} \qquad (15-5) \qquad . \ . \ .$$

وحيث ان β_j^* خالية من الوحدات لذا يتم استخدامها لإغراض المقارنة . فإذا كانت β_1^* ضعف β_2^* فذلك يدلل على ان المتغير X_2 هو ضعف أهمية المتغير X_1 في تقدير قيمة Y.

مثال(11−5) : وبالرجوع الى المثال :(8−5) فان معلمات الانحدار القياسية :

$$\hat{\beta}_j^* = \hat{\beta}_j \cdot \sqrt{\frac{S_{X_j X_j}}{S_{YY}}}$$

$$\hat{\beta}_1^* = (0.75) \cdot \sqrt{\frac{5.6}{61}} = 0.227$$

$$\hat{\beta}_2^* = (3)\cdot\sqrt{\frac{8.4}{61}} = 1.1132$$

$$\hat{\beta}_3^* = (1.5)\cdot\sqrt{\frac{2.4}{61}} = 0.298$$

وعلية فان المتغير X_2 هو أكثر المتغيرات أهمية في تفسير قيم Yويليه المتغير X_3 في الأهمية، ثم المتغير X_1 هو اقلها أهمية .

مثال(12−5) : معادلة تقدير النموذج بصيغته القياسية : $\hat{Y}_t^* = 0.069\ X_1^* + 0.823\ X_2^*$

وحيث أن $\hat{\beta}_1^* = 0.069$ وهي أقل من $\hat{\beta}_2^* = 0.823$

لذلك فأن المتغير X_2 له أهمية (12) مرة أكبر من المتغير X_1.

مثال(13−5) : عينة افتراضية بحجم (n = 10) مشاهدات ، صيغتها التقديرية

$$\hat{y} = 0.432x_1 - 0.82x_2 + 0.789x_3$$

$$\bar{Y} = 5 \quad , \quad \bar{X}_3 = 1 \quad , \quad \bar{X}_2 = 2 \quad \bar{X}_1 = 0 \quad , \quad TSS = 156$$

ومصفوفة العزم الثاني لانحرافاتY، X_1، X_2، X_3 عن متوسطاتها :

$$(x'x : x'y) = \left[\begin{array}{ccc|c} 60 & -53 & 34 & 91 \\ & 48 & -31 & -82 \\ & & 24 & 56 \end{array}\right]$$

م/ (1) احسب الصيغة التقديرية القياسية :

(2) اختبر أهمية أثر X_3 الكلي والإضافي .

(3) احسب الأهميةالإضافية للمتغير X_3 على النموذج :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

(4) اختبر الفرضية :

$$H_0 = (0 \quad 1 \quad 1)\begin{pmatrix}\beta_1 \\ \beta_2 \\ \beta_3\end{pmatrix} = \begin{pmatrix}0 \\ 0 \\ 0\end{pmatrix}$$

(5) اختبر الفرضية : $H_0 = R\beta = 0$ علماً بان :

$$R = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

الجواب:(1) الصيغة التقديرية القياسية ، بالاعتماد على العلاقة (4−7) :

$$\hat{\beta}_j^* = \hat{\beta}_j \cdot \sqrt{\frac{\Sigma x_j^2}{\Sigma y_j^2}}$$

$$\hat{\beta}_1^* = (0.432) \cdot \sqrt{\frac{60}{156}} = 0.2679$$

$$\hat{\beta}_2^* = (-0.82) \cdot \sqrt{\frac{48}{156}} = -0.455$$

$$\hat{\beta}_3^* = (0.789) \cdot \sqrt{\frac{24}{156}} = 0.309$$

وعلية فان أهمية المتغيرات التوضيحية بالتتابع هي : X_2 , X_3 ثم X_1 .

$$\hat{y}^* = 0.2679 x_i^* - 0.455 x_2^* + 0.309 x_3^*$$

(2) الأثر الكلي لـ X_3 : $H_0 : \beta_1 \neq 0$ $vs.$ $H_0 : \beta_3 = 0$

هي قيمة المعلمة باستخدام الانحدار البسيط

$$t^*_{\hat{\beta}_3} = \frac{\hat{\beta}_3}{s.e(\hat{\beta}_3)}$$

$$\hat{\beta}_3 = \frac{\Sigma x_3 y}{\Sigma x_3^2} = 2.333$$

$$ESS(X_3) = \frac{(\Sigma x_3 y)^2}{\Sigma x_3^2} = \frac{(56)^2}{24} = 130.7$$

$$RSS(X_3) = TSS - ESS(X_3) = 156 - 130.7 = 25.3$$

$$\frac{RSS(X_3)}{n-2} = \frac{25.3}{8} = 3.16$$

$$s.e(\hat{\beta}_3) = \sqrt{3.16 / \Sigma x_3^2} = 0.36$$

$$t^*_{\hat{\beta}_3} = \frac{2.333}{0.36} = 6.48$$

تقارن مع القيمة الجدولية $t_{c(8,0.025)} = 2.306$

القرار : X_3 مهم معنوياً لتحديد التغيرات في Y .

لاختبار الأثرالإضافيالذييضيفه X_3 للنموذج المتضمن X_1 و X_2فان فرضية العدم :

$$H_0 : \beta_3 / X_1 X_2 = 0 \qquad vs. \qquad H_1 : \beta_3 / X_1 X_2 \; const \neq 0$$

$$ESS\ (X_3/X_1X_2\ ) = ESS(X_1X_2X_3) - ESS(X_1X_2)$$

$$ESS(X_1X_1X_3) = (0.432)(91) + (-0.82)(-82) + (0.789)(56)$$
$$= 39.312 + 67.24 + 44.184 = 150.7$$

لحساب $ESS(X_1\ , X_2)$ ، لابد من حساب $\hat{\beta}_2 , \hat{\beta}_1$ أولا

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{bmatrix} 60 & -53 \\ & 48 \end{bmatrix}^{-1} \begin{pmatrix} 91 \\ -82 \end{pmatrix}$$

$$(x'x)^{-1} = \begin{bmatrix} 0.676 & 0.746 \\ & 0.845 \end{bmatrix}$$

المحدد =(2880−2809) =71

$$\therefore \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{bmatrix} 0.676 & 0.746 \\ 0.746 & 0.845 \end{bmatrix} \begin{pmatrix} 91 \\ -82 \end{pmatrix} = \begin{pmatrix} 0.344 \\ -1.404 \end{pmatrix}$$

أذن : $ESS(X_1X_2) = \hat{\beta} \acute{x}' y$

$$ESS(X_1X_2\ ) = (0.344 \quad -1.404) \begin{pmatrix} 91 \\ -82 \end{pmatrix} = 31.304 + 115.182 = 146.432$$

$$ESS(X_3/X_1X_2) = 150.7 - 146.432 = 4.268$$

$$RSS(X_1X_2X_3) = TSS - ESS(X_1X_2X_3) = 156 - 150.7 = 5.3$$

$$\frac{RSS(X_1X_2X_3)}{n-k-1} = \frac{5.3}{6} = 0.883$$

$$\Rightarrow F^* = \frac{4.268}{0.883} = 4.834$$

تقارن مع $F_{(1,6,0.95)} = 5.99$

اذن القرار لا نرفض فرضية العدم .

أيأنX_3 غير مهم في تحديد قيمة Y .

(3)الأهميةالإضافية للمتغير X_3 :

$$r^2_{YX_3 \,.\, X_1X_2} = \frac{ESS(X_3/X_1X_2)}{RSS(X_1X_2)}$$

$$RSS(X_1X_2) = TSS - ESS(X_1X_2)$$

$$= 156 - 146.432 = 9.568$$

$$r^2_{YX_3 \,.\, X_1X_2} = \frac{4.268}{9.568} = 0.446$$

(٤) اختبر :

$$H_0:[0 \quad 1 \quad 1]\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

أي:

$$H_0:\beta_2+\beta_3=0 \qquad VS \qquad H_1:\beta_2+\beta_3\neq 0$$

هناك طريقتان لإجراء الاختبار. الطريقة الأولى باستخدام الفرضية الخطية العامة وعلى وفق آلاتي:

النموذج الكلي (غير المقيد)

النموذج المقيد : بتعويض : $\beta_2 = -\beta_3$

$$y = \beta_1 x_1 + (-\beta_3)x_2 + \beta_3 x_3 + u$$

$let \quad z_2 = x_3 - x_2 \quad , \quad x_1 = z_1$ **Error! Bookmark not defined.**

$$y = \beta_1 \underbrace{x_1}_{z_1} + \beta_3 \underbrace{(x_3 - x_2)}_{z_2} + u$$

فالنموذج المقيد :

$$y = \alpha_1 z_1 + \alpha_2 z_2 + v$$

$$\begin{pmatrix}\hat{\alpha}_1\\ \hat{\alpha}_2\end{pmatrix} = \begin{bmatrix}\Sigma x_1^2 & \Sigma x_1(x_3-x_2)\\ & \Sigma(x_3-x_2)^2\end{bmatrix}^{-1} x'y$$

$$= \begin{bmatrix}\sum x_1^2 & \sum x_1x_3 - \sum x_1x_2\\ & \sum x_3^2 + \sum x_2^2 - 2\sum x_2x_3\end{bmatrix}^{-1} \begin{bmatrix}\sum x_1 y\\ \sum x_3 y - \sum x_2 y\end{bmatrix}$$

$$= \begin{bmatrix}60 & 34-(-53)\\ & 24+48-2(-31)\end{bmatrix}^{-1}\begin{bmatrix}91\\ 56-(-82)\end{bmatrix} = \begin{bmatrix}60 & 87\\ 87 & 134\end{bmatrix}^{-1}\begin{bmatrix}91\\138\end{bmatrix}$$

المحدد = $8040 - 7569 = 471$

$$(x'x)^{-1}_{z_1z_2} = \begin{bmatrix}0.2845 & -0.185\\ & 0.127\end{bmatrix}$$

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{bmatrix} 0.2845 & -0.185 \\ & 0.127 \end{bmatrix} \begin{bmatrix} 91 \\ 138 \end{bmatrix} = \begin{bmatrix} 25.889 - 25.53 \\ -16.835 + 17.526 \end{bmatrix} = \begin{bmatrix} 0.3594 \\ 0.691 \end{bmatrix}$$

مجموع المربعات المشروحة للنموذج المقيد :

$$ESS_r = [\hat{\alpha}_1 \quad \hat{\alpha}_2] \begin{pmatrix} \Sigma z_1 Y \\ \Sigma z_2 Y \end{pmatrix} = [0.3594 \quad 0.691] \begin{bmatrix} 91 \\ 138 \end{bmatrix} = 128.0634$$

وبتطبيق العلاقة (5-13)

$$F^* = \frac{(ESS_U - ESS_r)/1}{RSS_U/(n-k-1)} = \frac{(150.7 - 128.0634)}{0.885} = 25.578$$

وبمقارنتها مع ($F_{(1,6,0.95)} = 5.99$) فان القرار يكون برفض H_0 . أي أن القيد يحسن القوة التنبؤية.

أما الطريقة الثانية :

بما أن الفرضية $H_0 : \beta_2 + \beta_3 = 0$ ، فهي تركيب خطي بدلالة المعلمات β_2 و β_3 وبذلك يمكن الاختبار باستخدام الاحصاءة t وكالاتي:

$$t^*_{\hat{\beta}_2 + \hat{\beta}_3} = \frac{\hat{\beta}_2 + \hat{\beta}_3}{s.e(\hat{\beta}_2 + \hat{\beta}_3)}$$

$$s.e(\hat{\beta}_2 + \hat{\beta}_3) = \sqrt{\hat{\sigma}^2 c_{22} + \hat{\sigma}^2 c_{33} + 2\hat{\sigma}^2 c_{23}}$$

غير أن الانحراف المعياري للتركيب يستوجب معرفة معكوس المصفوفة $(X'X)$

$$(X'X) = \begin{bmatrix} 60 & -53 & 34 \\ & 48 & -31 \\ & & 24 \end{bmatrix}$$

في جميع أمثلة الفصلين (الرابع و الخامس) تم استخدام الطرائق الحسابية الاعتيادية في الحلول، ويمكن استخدام البرامج الجاهزة لهذا الغرض وسيتم استخدام برنامج SPSS على بيانات المثال (3-4) بعد إضافة المتغير(X_2) الذي يمثل درجة حرارة سائل تبريد الماكنة الى بيانات المثال .

مثال(٥ – ١٤): اضافة المتغير(X_2) الذي يمثل درجة حرارة سائل تبريد الماكنة الى بيانات المثال (3-4)، وكما هو معروض في الجدول(٥– ٣)

الجدول (٥- ٣)

عدد العيوب (Y) وسنوات الخدمة (X_1) ودرجة حرارة سائل التبريد (X_2) لعينة من (15) مشاهدة سحبت عشوائياً من مصنع معين.

Y	39	٣٥	٤٠	28	35	30	33	33	26	40	41	33	٣١	33	38
X_1	5	11	4	12	11	13	9	11	12	4	5	6	٧	٩	٦
X_2	36	35	31	30	37	32	36	34	37	30	31	32	33	34	35

المصدر: http // samehar. word press.com

وبعد ادخال البيانات واستخدام البرنامج spss تظهر نتائج التقدير:

جدول (5 - 4)

نتائج التحليل الخاصة بالمعلمات

Model	Unstandardized Coefficients		Standardized Coefficients
	β (1)	Std. Error (2)	Beta (3)
(Constant)	39.708	10.891	
X_1	-1.175	0.260	-.824
X_2	0.132	0.340	0.071

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

والتي يمكن عرضها كالاتي:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

$$\hat{Y} = 39.708 - 1.175 X_1 + 0.132 X_2$$

ويفسر معلمة الانحدار الخاص بالمتغير المستقل (X_1) ان زيادة خبرة العامل بمقدار سنة واحدة سيؤدي الى انخفاض في عدد العيوب بمقدار (1.175) ، وعند زيادة درجة حرارة سائل تبريد الماكنة (X_2) بمقدار درجة سيليزية واحدة تؤدي الى ارتفاع عيوب المنتج بمقدار (0.132) .

فالعمود (1) يمثل قيم المعلمات المقدرة غير القياسية.

العمود (2) يمثل الخطأ المعياري للمعلمات المقدرة غير القياسية.

أما العمود(3) يمثل المعلمات المقدرة القياسية (المعيارية).

وتظهر القيم ان المتغير X_1 هو أكثر تأثيراً فى Y من المتغير X_2 على وفق قيمته القياسية المقدرة.

والجدول(٥–٥) يعرض معلومات تفيد لاختبار معنوية المعلمات على وفق الاحصاءة t الى جانب مجال الثقة باحتمال 95% وكالاتي:

جدول (5-5)

نتائج التحليل الخاصة بالمعلمات

Model	t (1)	Sig.(p) (2)	95% confidence interval for B	
			Lower Bound (3)	Upper Bound (4)
(Constant)	3.646	0.003	15.978	63.438
X_1	-4.527	0.001	-1.741	-0.610
X_2	0.388	0.705	-0.608	0.872

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

فالعمود (1) يعرض قيم الاحصاءة t للمعلمات المقدرة والتي تسمح لاختبار الفرضيات:

$H_0: \beta_i = 0 \qquad i = 0, 1, 2$

vs.

$H_1: \beta_i \neq 0$

$$\left| t^*_{\hat{\beta}_i} \right| = \frac{\hat{\beta}_i}{s.e(\hat{\beta}_i)}$$

فان $\left| t^*_{\hat{\beta}_0} \right| = (3.646)$

وكذلك $\left| t^*_{\hat{\beta}_1} \right| = (4.527)$

وكذلك $\left| t^*_{\hat{\beta}_2} \right| = (0.388)$

والعمود (2) يعرض قيم p آزاء كل معلمة ، فنلاحظ ان قيمة p اقل من ٥% $(p < 0.05)$ بالنسبة للمتغير المسقل X_1 وهذا يعني ان المعلمة $\hat{\beta}_1$ معنوية احصائياً باستخدام مستوى دلالة ٥%، أي ان خبرة العاملمهمة في تفسيرالتغيرات في جودة المنتج. كما نلاحظ ان قيمة p اكبر من ٥%$(p > 0.05)$ بالنسبة للمتغير المسقل X_2 وهذا يعني ان المعلمة $\hat{\beta}_2$ غير معنوية احصائياً أي ان درجة حرارة سائل التبريد غير مهمة في تفسير التغيرات في عيوب المنتج.

اما العمودان (3) و (4) فتظهر مجال الثقة باحتمال 5% للمعلمات المقدرة بحديها الادنى والاعلى.

$$\hat{\beta}_i - s.e(\hat{\beta}_i).(t_{c(n-k-1,0.025)}) \langle \ \hat{\beta}_i \ \langle \ \hat{\beta}_i + s.e(\hat{\beta}_i).(t_{c(n-k-1,0.025)})$$

أي ان مجال الثقة باحتمال 5% للمعلمات المقدرة يمكن تثبيته من الجدول كالاتي:

لـ $\hat{\beta}_0$ $15.978 \ \langle \ \hat{\beta}_0 \ \langle \ 63.438$

لـ $\hat{\beta}_1$ $-1.741 \ \langle \ \hat{\beta}_1 \ \langle \ -0.610$

لـ $\hat{\beta}_2$ $-0.608 \ \langle \ \hat{\beta}_2 \ \langle \ 0.872$

وان مخرجات البرنامج توفر جدول تحليل التباين والذي يساعد في اختبار معنوية النموذج ككل على وفق اختبار F وكالاتي:

جدول (6-5)

نتائج تحليل التباين

ANOVA

S.O.V	SS	d.f	MSS	F	Sig
Regression	188.179	2	94.089	10.945	.002
Residual	103.155	12	8.596		
Total	291.333	14			

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

ويوضح الجدول (6-5) نتائج تحليل تباين الانحدار(ANOVA) فيستدل من خلاله على نسبة التباين الذي تفسره المتغيرات المستقلة (خبرة العامل) و (درجة حرارة سائل التبريد) من تباين المتغير التابع (جودة المنتج)، وبما ان مستوى الدلالة هو (P = 0.002) وهو اقل من (0.05) وبالتالي فان قيمة (F) دالة إحصائياً، اي يمكننا القول ان النموذج يمثل البيانات خير تمثيل عند مستوى دلالة اقل من (0.05).

وتكون مخرجات البرنامج بالنسبة لمعامل التحديد كما في الجدول (7-5)

جدول (7-5)

ملخص نتائج تحليل الانحدار

R	R Square (R^2)	Adjusted R Square ($\overline{R}^2$)	Std. Error of the Estimate
.804	.646	0.587	2.93

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

نلاحظ في جدول (5-7) معامل الارتباط المتعددR) (0.804 = والذي يبين مدى تأثير المتغيرين المستقلين في المتغير التابع أي ان هناك ارتباطاً قوياً بين المتغيرات المستقلة والمتغير التابع في النموذج. كما يظهر في الجدول أيضا معامل التحديد الذي يفسر نتائج النموذج، وقيمته ($R^2 = 0.646$) وتعني ان النموذج المقدر يعبر عن أكثر من(64)% من البيانات أي ان المتغيراتالمستقلة تفسر أكثر من(64) % من تباين المتغير التابع، وتم حساب معامل التحديد المعدل للنموذج والذي يساوي ($\bar{R}^2 = 0.587$) ويستخدم للغرض نفسه اعلاه ولكن بصورة ادق.

أسئلة الفصل الخامس

س1:(أ) اشتق صيغة ملائمة لحساب مصفوفة التباين والتباين المشترك للمعلمة β_0 بافتراض النموذج: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$.

(ب) اختبر معنوية المعلمة β_0 من واقع المعلومات: $\hat{Y} = -5.3 + 2.5X_1 + 0.78X_2$

$$(X'X)^{-1} = \begin{pmatrix} 0.5 & 0.1 \\ & 0.3 \end{pmatrix} \quad , \quad RSS = 8.27 \quad , \quad n = 20 \quad , \quad \Sigma X_1 = 100 \quad , \quad \Sigma X_2 = 120$$

(ج) تنبأ عن قيمة Y عندما $X_1 = 10$ ، $X_2 = 10$ ثم احسب مجال الثقة للقيمة التنبؤية الجديدة.

س2: مع توافر لوحة البيانات التالية:

s.o.v	d.f	ss
$R(X_1X_2X_3/X_0)$	3	678.294
$R(X_1X_2/X_3X_0)$	2	15.6
$R(X_1X_3/ X_2X_0)$	2	
$R(X_2/ X_0)$	1	662.7
$R(X_3/ X_0)$	1	
$Error(X_1X_2X_3$	5	0.056
Total	8	678.35

اختبر الفرضيات التالية:

١. $H_0 : R\beta=٠$ حيث أن :

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

٢. الأثر المباشر للمتغير X_3.

٣. الأثر الإضافي للمتغيرين X_3 ، X_1 علماً بوجود المتغير X_2.

س3: في نموذج الانحدار الخطي المتعدد كيف تعبر إحصائيا عما يلي:

١. عدم وجود علاقة خطية تامة بين المتغيرات التوضيحية.

٢. خاصية تجانس خطأ التقدير.

٣. الاستقلال الذاتي لخطأ التقدير.

س4: عينة من (30) مشاهدة Y، X_1، X_2، X_3 ولدت لوحة البيانات التالية:

$$(X'X)^{-1}=\begin{pmatrix}7.96 & -0.39 & -0.04 & -0.62\\ & 0.05 & -0.02 & 0.06\\ & & 0.05 & 0.01\\ & & & 0.11\end{pmatrix}\quad,\quad X'Y=\begin{pmatrix}120\\437\\254\\467\end{pmatrix}\quad,\quad Y'Y=1800$$

١. افترض تحقق فروض التحليل، هل نتائج التقدير باستخدام OLS تختلف عنها عند استخدام طريقة الإمكان الأعظم.

٢. احسب معادلة التقدير وفسر نتائج التقدير.

٣. اختبر الأهمية الإضافية للمتغيرين X_2 و X_3 للنموذج: $Y=\beta_0+\beta_1X+u_i$.

٤. احسب الأهمية النسبية لإضافة المتغيرين X_2 و X_3 .

٥. احسب القيمة التنبؤية لـ Y عندما $X_1 = 2$ و $X_2 = 5$ و $X_3 = -2$ واحسب مجال الثقة للقيمة التنبؤية باحتمال 95%.

س٥: مع توافر لوحة المعلومات لعينة (n = 14)

s.o.v	$R(X_1X_2X_3/X_0)$	$R(X_1X_2/X_0)$	$R(X_1X_3/X_0)$	$R(X_2X_3/X_0)$	$R(X_1X_0)$	$R(X_2/X_0)$	$R(X_3/X_0)$
SS	725	414	645	348	325	242	60

Total = 1325

١. اختبر الأثر المباشر للمتغير X_2.

٢. اختبر الأثر الإضافي للمتغير X_3 على معادلة الانحدار:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+u_i$$

٣. $H_0 : R\beta=$٠ علماً أن:

$$R=\begin{pmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

س6: عينة من (40) مشاهدة أعطت النتائج التالية:

$$\ln\hat{Y}=1.72-0.60\ln X_1+0.351\ln X_2 \qquad , \qquad r_{X_1X_2}=0.31$$

$$s.e\ :\qquad (0.326)\qquad\qquad (0.081)$$

أختبر:

١. تساوي المرونات بالنسبة للعنصرين X_1 و X_2 .

٢. مجموع المرونات تساوي واحداً صحيحاً.

س7: مع توافر المعلومات:

$$(X'X)^{-1} = \begin{pmatrix} 9.1 & -0.5 & -0.8 & -2.3 \\ & 0.25 & 0 & 0 \\ & & 0.16 & 0.2 \\ & & & 0.7 \end{pmatrix} \quad , \quad \sum_{i=1}^{20} e_i^2 = 3.67$$

$$\hat{Y} = -0.6 + 0.2X_2 + 3X_3$$

١. احسب مجال الثقة باحتمال 95% للتركيب ($\beta_2 - \beta_3$).

٢. مجموع المربعات الإضافية للمتغيرين X_2 و X_3.

٣. الأهمية النسبية للتغيرات المشروحة من قبل المتغير X_1 فقط في النموذج.

س8: من واقع البيانات:

$$\hat{Y} = -5.3 + 2.5X_1 - 0.78X_2 \qquad i = 1,\cdots,15$$

$$RSS = 8.27 \quad , \quad (X'X)^{-1} = \begin{pmatrix} 5.6 & -0.8 \\ & 8.4 \end{pmatrix} \quad , \quad \Sigma X_1 = 100 \quad , \quad \Sigma X_2 = 120$$

١. فسر المدلول الإحصائي للمعلمة β_0.

٢. أختبر معنوية المعلمة β_0.

س9: ما الشروط الواجب توافرها لأجل تساوي المعلمات المقدرة باستخدام صيغة النموذج المتعدد مع نظيراتها من النموذج الخطي البسيط. وضح حالة الانحدار بثلاثة متغيرات Y و X_1 و X_2 .

س10: مع توافر المعلومات التالية:

$$(X'X) = \begin{pmatrix} 12 & 8 \\ & 12 \end{pmatrix} \quad , \quad \Sigma x'y = \begin{pmatrix} 10 \\ 8 \end{pmatrix} \quad , \quad \bar{Y} = 20 \quad , \quad \bar{X}_1 = 10 \quad , \quad \bar{X}_2 = 5 \quad , \quad \Sigma y^2 = 10$$

١. احسب الانحراف المعياري للمقدرات.

٢. تنبأ عن قيمة Y عندما $2X_2$ = وX_1 10= .

٣. احسب مجال الثقة باحتمال 95% للقيمة التنبؤية لـ Y.

٤. احسب معامل التحديد ومعامل التحديد المعدل وفسر دلالته.

س11: صحح الخطأ إن وجد.

١. رتبة المصفوفة X تكون تامة من ناحية الأعمدة إذا انعدم وجود ارتباط خطي تام بين المتغيرات التوضيحية.

٢. في الانحدار $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$ يكون $\text{var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\Sigma x_2^2 (1 - r_{12}^2)}$

٣. تفضل العلاقة $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$ مع $R^2 = 0.8$ على العلاقة $\hat{Z} = \hat{\alpha}_0 + \hat{\alpha}_1 \ln X_1 + \hat{\alpha}_2 \ln X_2$ مع $R^2 = 0.7$ إذا كان حجم العينة 12 في العلاقتين متساوياً.

٤. إن أثر المتغير X على المتغير Y يتحدد بمقدار β_1إذا كانت معادلة التقدير $\ln \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$.

٥. إذا كانت فترة الثقة للقيمة التنبؤية باحتمال 99% هي ($0.3 \le Y_{of} \le 7.8$) فان القيمة التنبؤية تعد غير معنوية باستخدام مستوى معنوية 1%.

٦. إذا $r_{12} = 0$ و $r_{13} \neq 0$ و $r_{23} \neq 0$ فان $r_{12.3} > 0$.

٧. رفض الفرضية: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ دليل على إن جميع المتغيرات المستقلة مهمة لتحديد استجابة Y.

٨. يتساوى الأثر المباشر مع الأثر الإجمالي في معادلات تحليل الانحدار بشكل عام،

الفصل السادس

اختلال فروض التحليل التي تخص توصيف النموذج وحد الاضطراب العشوائي

(6−1)اختلال الفرضية (3).

الفرضية (3): " متوسط المتغير العشوائي يساوي صفراً.

في النموذج الخطي المتعدد:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \quad , \quad t=1,2, ... , n$$

ان اختلال الفرض $E(u_i) = 0$ يكمن في احتمالين:

(أ) $$E(u_t / X_1 X_2, \cdots, X_k) = w = constant \neq 0$$

وبذلك فان معادلة الانحدار:

$$E(Y_t / X_1, X_2, \cdots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + w$$
$$= (\beta_0 + w) + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

وبذلك فان اختلال الفرض يؤدي إلى ان تكون قيمة المقطع الصادي $\alpha = (\beta_0 + w)$.

أي ان المقطع الصادي المقدر بطريقة OLS يكون متحيزاً أي: $E(\hat{\alpha}) = E(\beta_0 + w) \neq \beta$

وحيث ان المهم في الدراسات التطبيقية هو تقدير معلمات الانحدار، وعليه فلا توجد مشكلة في هذه الحالة.

(ب) اذا $$E(u_t / X_1 X_2, \cdots, X_k) = w_t \quad \forall \quad t = 1,2, \cdots, n$$

في هذه الحالة لا يمكن تقدير معلمات الانحدار بطريقة OLS وذلك لأن عدد المعلمات المطلوب تقديرها (n+k−1) سيكون أكبر من عدد المشاهدات (n) .

(6−2) اختلال الفرضية (11).

الفرضية (11): التوزيع الطبيعي لـ U"u " Normality assumption of

اذا كان الهدف الرئيس هو التقدير فقط ، فان افتراض المتغير العشوائي يتوزع توزيعاً طبيعياً ليس بالفرضية الأساسية. أي ان عدم تحقق فرضية التوزيع الطبيعي لـ u لا يفقد المقدرات صفة (BLUE).

غير ان المقدرات لاتمتلك صفة التوزيع الطبيعي في العينات الصغيرة وكذلك تستمر الصفة مع كبر حجم العينة. لذا فان استخدام الاختبارات المختلفة F و t ممكناً في حالة العينات الكبيرة، أما مع حجم العينة

المحدود أو الصغير فلا تتحقق صفة التوزيع الطبيعي. ومعلوم إحصائياً إن فرضية التوزيع الطبيعي تكون مهمة إذا كان الهدف هو لاختبار الفرضيات.

وعليه فانه مع العمل بحجم محدود أو صغير من المشاهدات يتطلب إجراء اختبار للتحقق من التوزيع الطبيعي. ومن الاختبارات المستخدمة في هذا الشأن هو اختبار Jarque – Bera test والذي يعتمد على اختبار الفرضية:

$$H_0 : s = 0 \quad ; \quad k = 3$$

$$vs \quad H_0 : s \neq 0 \quad ; \quad k \neq 3$$

اذ ان التوزيع الطبيعي يكون متماثلاً او ان معامل الانحراف $s = 0$

أي ان الانحراف عن المتوسط مساوياً الى الصفر. الى جانب ان معامل التفلطح (kurtosis) للتوزيع الطبيعي هو (k = 3) .

علماً بان s& k لاي متغير عشوائي وليكن Y، تحسب على وفق الاتي:

$$s = \frac{\text{مربع العزم الثالث حول المتوسط}}{\text{مكعب العزم الثاني}} = \frac{\left[\mathrm{E}(Y-\mu)^3\right]^2}{\left[\mathrm{E}(Y-\mu)^2\right]^3} \quad \cdots \quad (1-6)$$

$$k = \frac{\text{العزم الرابع}}{\text{مربع العزم الثاني}} = \frac{\mathrm{E}(Y-\mu)^4}{\left[\mathrm{E}(Y-\mu)^2\right]^2} \quad \cdots \quad (2-6)$$

أما المختبر المقترح فهو:

$$J.B = n\left(\frac{s^2}{6} + \frac{(k-3)^2}{24}\right) \quad \cdots \quad (3-6)$$

وعليه لاختبار التوزيع الطبيعي للمتغير العشوائي u يتم حساب : s و k لبواقي علاقة الانحدار (ei) ثم يطبق المختبر J.B ، علماً أن المختبر (J.B) يتوزع توزيع مربع كاي : $J.B \sim \chi_2^2$ ، وبدرجات حرية مقدارها (2) .

كما ويمكن بشكل عام استخدام المدرج التكراري للبواقي (Histogram) ومن خلال المدرج يستدل فيما اذا كان يشبه التوزيع الطبيعي أم لا .

أما الاختبار الآخر فيكمن بمقارنة الانحراف المعياري للانحدار$(\hat{\sigma})$ مع مدى قيم البواقي مقسومة على(6) فاذا كان: ($s \cong \frac{\text{مدى } \hat{e}}{6}$) فان التوزيع الطبيعي يتحقق.

علماً بان المدى هو (قيمة e الأكبر – القيمة الأصغر).

مثال (6-1): في أدناه بواقي علاقة انحدار ل (١٥) مشاهدة:

e	−2.857	−4.68	3.807	1.984	0.161	3.338	−3.33	5.312	−1.821
i	1	2	3	4	5	6	7	8	9

↓

e	−3.489	−0.157	7.865	1.352	−0.471	−6.984
i	10	11	12	13	14	15

$\Sigma e_t^2 = 227.1668$

$$\hat{\sigma}^2 = 17.47$$

$$\hat{\sigma} = 4.1797$$

$$\hat{\sigma} = s$$

$$\Leftarrow \quad \frac{\text{مدى}}{6} = 2.47$$

إما باستخدام اختبار Jarqa – Bera فيمكن تطبيق القوانين (٦-١) و(٦-٢) و(٦-٣) تباعا كما بالإمكان اعتماد البرنامج الجاهز(SPSS) لحساب الأحصاءة (J–B):

s = 0.230483

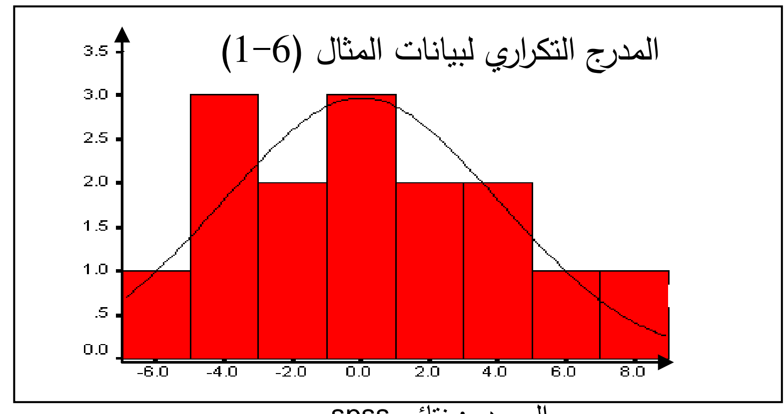
k = 2.415689 و

$J.B = 0.346193 \qquad \chi^2_{2(0.05)} = 7.81473$

القرار نقبل فرضية العدم ، أي لانرفض ان التوزيع طبيعي.

كما يمكن الاستدلال من خلال رسم (Histogram)

وعند رسم (Histogram) لبيانات المثال اعتمادا على (SPSS) يتضح بان التوزيع الطبيعي يمكن عدّه مناسبا للبيانات والشكل التالي يعكس ذلك



المصدر : نتائج spss

مثال (6-2): توفرت اخطاء علاقة انحدار Y على X_1 و X_2 على وفق الآتي:

e	5.777	-4.746	9.848	2.563	-0.746	4.07702	-1.426	22.82884
i	1	2	3	4	5	6	7	8

↓

e	-1.42811	-6.8166	38.85023	-68.7819
i	٩	10	11	12

اختبر ان البواقي تتوزع توزيعاً طبيعياً ؟ $n=12$, $\Sigma e^2 = 6988.634$

الحل:

نستخدم اختبار Jarqua – Bera

$S = -1.506426$

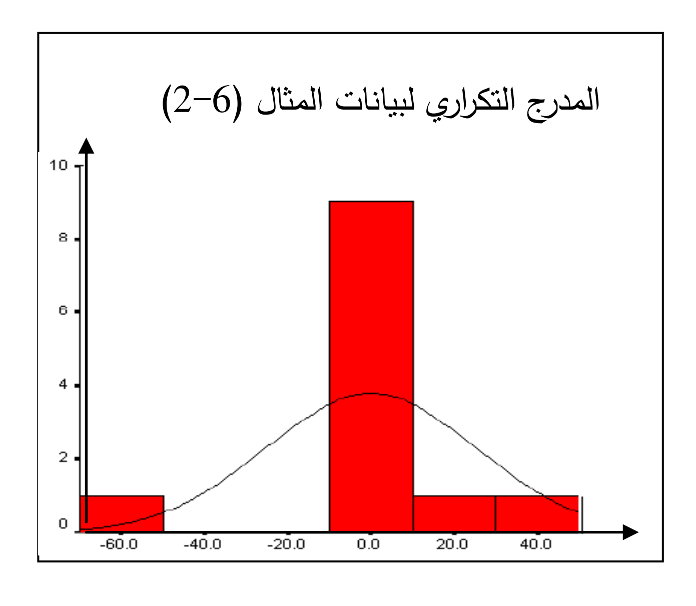
$k = 6.128895$

$\therefore \quad J.B = 9.433629$

نقارن مع الجدولية $\chi^2_{2(0.05)} = 7.81473$

اذن نرفض فرضية العدم ، أي ان التوزيع غير طبيعي.

وعند رسم Histogram بمساعدة برنامج (SPSS) يعزز الاختبار:



المصدر : نتائج spss

يتضح أن التوزيع غير طبيعي وهو ما يؤكد ما توصلنا إليه باستخدام اختبار(Jarque – Bera test)

وبالنظر لمدى e : القيمة الأكبر = 38.85023 ، القيمة الأصغر = 68.7819 −

$\frac{\text{مدى}}{6} = 17.9$ ، se = 776.5 وعليه الفرضية للتوزيع الطبيعي لا تتحقق.

مثال(٦-٣) استخدم بيانات المثال (٥-١٤) نفسها لتوضيح استخدام البرنامج الجاهز (SPSS)

لاختبار جاركا- بيرا (Jarque – Bera (JB)Test)

ولمتابعة تحليل النتائج من خلال اختبار فرضيات التحليل يمكن اعتماد برنامج SPSS وباستخدام المختبر الاحصائي (Jarque – Bera (JB لاختبار تحقق فرض التوزيع الطبيعي .

$$H_0 : s = 0 \quad , \quad k = 3$$ التوزيع طبيعي

$$H_1 : s \neq 0 \quad , \quad k \neq 3$$ التوزيع غير طبيعي

$$s = \frac{\left[E(Y-\mu)^3\right]^2}{\left[E(Y-\mu)^2\right]^3} \quad , \quad k = \frac{E(Y-\mu)^4}{\left[E(Y-\mu)^2\right]^2}$$

$$J \cdot B = n\left[\frac{s^2}{6} + \frac{(k-3)^2}{24}\right] \sim \chi_2^2$$

باستخدام البرنامج يمكن حساب القيمة المقدرة لـ $\hat{Y}$ منها تحسب e_i حيث ان $e_i = Y_i - \hat{Y}_i$

- ثم نحسب $(e_i - \bar{e})^2$ و $(e_i - \bar{e})^3$ و $(e_i - \bar{e})^4$ ثم نحسب معامل التماثل s

$$s = \frac{(-141.64)^2}{(103.09)^3} = 0.02$$

ومعامل التفلطحk :

$$k = \frac{(1576.93)}{(103.09)^2} = 0.15$$

ثم نحسب معامل (J.B):

$$J.B = 15\left[\frac{(0.02)^2}{6} + \frac{(0.15-3)^2}{24}\right]$$

$$= 15\left[0.00007 + 0.34\right]$$

$$= 5.08$$

وبمقارنتها مع القيمة الجدولية لمربع كاي: $\chi^2_{c(2,0.05)} = 5.99$

بما إن قيمة J.B الحسابية أقل من χ^2 الجدولية نقبل فرضية العدم وهذا لا يعني بالضرورة ان التوزيع طبيعي وإنما رفض الفرضية القائلة $s \neq 0 \;,\; k \neq 3$.

أما كيفية تجاوز مشكلة إن بواقي العلاقة يتوزع توزيعاً غير طبيعي فيمكن توسيع حجم العينة أو تضمين متغيرات توضيحية إضافية إلى النموذج كما يمكن أن نغير الصيغة الدالية المستخدمة فبدل ان نستخدم الصيغة الخطية يتم استخدام الصيغة اللوغاريتمية المزدوجة.

إن اختلال التوزيع الطبيعي للخطأ يتلازم مع عدم تحقق فرضية تجانس تباين الأخطاء والتي سيتم تناولها في الفقرة القادمة.

(3-6) فرضية تجانس التباين للمتغير العشوائي u Homoscedasticity

من الفرضيات المهمة في تحليل الانحدار هي ان تباين المتغير العشوائي (u_i) في المجتمع لمعادلة الانحدار يفترض ان يكون متجانساً لقيم المتغير (أو المتغيرات) التوضيحية X المعطاة.
وبالرموز

$$\operatorname{var}(u_i / X) = \sigma^2 \qquad \forall \qquad i = 1, \cdots .n$$

وتسمى هذه الفرضية تجانس تباين البواقي (Homoscedasticity) .
فعند اختلال (أو عدم تحقق)هذه الفرضية تبرز ما تسمى بمشكلة عدم تجانس التباين (Hetroscedasticity)

(1-3-6) طبيعة المشكلة Natur of the problem

عندما تختلف التغيرات في قيم المتغير العشوائي (u_i) مع قيم (X_i) فان تباين (Y_i) سوف يرتبط مع التغيرات في قيم (X_i) (بالزيادة أو النقصان) ، وبذلك يكون تباين (Y_i) مختلف وغير متساو .
وبصيغة الرموز:

$$E(u_i)^2 = \sigma_i^{\,2} \qquad i = 1, \cdots, n$$

$$E(u_1)^2 = E(u_2)^2 = \;\cdots\; E(u_n)^2 = \sigma_n^{\,2}$$

وبذلك فان مصفوفة التباين والتباين المشترك للمتغير العشوائي:

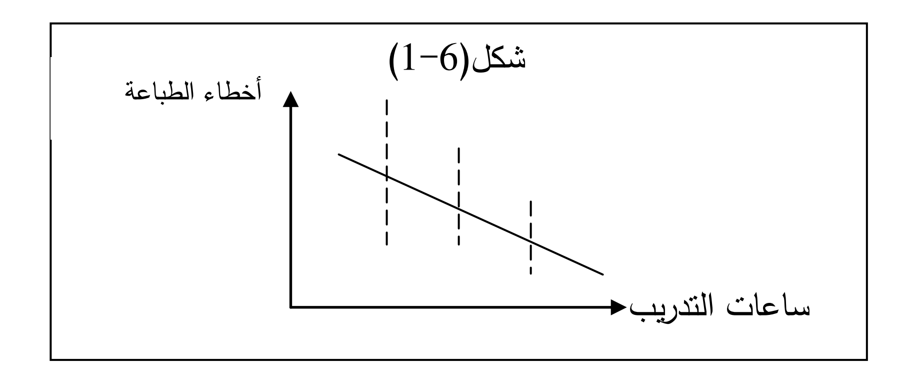
$$\text{var}(u_i) = \sigma^2\Omega$$

$$\Omega = \begin{pmatrix} \sigma_1^{\ 2} & 0 & \dots & 0 \\ 0 & \sigma_2^{\ 2} & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & & \sigma^2{}_n \end{pmatrix}$$

(6-3-2) أسباب المشكلة Causes of the problem

توجد في الواقع العملي أسباب مختلفة لوجود مشكلة عدم التجانس ، ويمكن التعرض الى أهمها:

١. اتباع نماذج أخطاء التعلم، ومع زيادة التعلم فان الأخطاء تنخفض وبذلك فان تشتت الأخطاء يتناقص لدى الأشخاص الذين تتزايد ساعات تدريبهم . وخير مثال على ذلك ان الأخطاء الطباعية تتناقص مع زيادة ساعات التدريب على الطباعة كما في الشكل التالي.



٢. طبيعة بعض الدراسات توحي الى أن تشتت بواقي العلاقة غير ثابتة. مثلا في دراسة دالة الادخار، فان تباين الادخار يتغير على وفق فئات الدخل المختلفة. ففي المتوسط ان الفئات الدخلية العالية تميل الى الادخار أكثر من فئات الدخل المنخفضة ، مع بقاء وجود تغيرات في الادخار للعوائل المختلفة. وبذلك فان انحدار الادخار كدالة بدلالة الدخل يولد تبايناً متزايداً مع الدخل وذلك لان الأفراد تكون لديهم خيارات أوسع حول سلوكهم الادخاري.
والحال نفسه عند دراسة الأرباح لعينة من الشركات فالشركات التي تكون أرباحهم كبيرة يتوقع ان تكون لديهم تغيرات واسعة حول سياساتهم الاستثمارية أكثر من الشركات ذات الأرباح المنخفضة.

٣. تحسين طرائق جمع البيانات بشكل عام يولد اتجاه حول انخفاض التباين.

٤. وجود المشاهدات الشاردة (الشاذة) سبباً لظهور عدم التجانس للأخطاء . فان حذف او تضمين مشاهدات شاذة تكون حافزاً للتأثير فى النتائج وخاصة في حالة العينة الصغيرة.

٥. سوء توصيف النموذج يعد مصدراً خصباً لوجود عدم التجانس. ويتولد سوء التوصيف في نموذج الانحدار من أمرين. الأول هو حذف متغير مهم أو إضافة متغير ليس بذي جدوى. أما الأمر الثاني فهو عدم استعمال الصيغة الدالية الملائمة للبيانات. فاستعمال الصيغية الخطية عوضاً عن الصيغة اللوغارتمية المزدوجة قد تكون سبباً لظهور مشكلة عدم التجانس. كما إن إجراء التحويلات غير المناسبة للبيانات تعد سبباً لحدوث مشكلة عدم التجانس.

لا بد من التنويه على ان مشكلة عدم التجانس تظهر غالبا في دراسات المقاطع العرضية (Cross–section) اذ يتم التعامل مع أفراد من المجتمع في لحظة زمنية معينة، وتكون هذه الأفراد غير متجانسة : مثل شركات كبيرة ومتوسطة وصغيرة، او فئات دخلية مختلفة. غير ان دراسات السلاسل الزمنية(Time–Series) ليست بمأمن من هذه المشكلة.

(3–3–6) تأثير عدم التجانس في نتائج طريقة المربعات الصغرى. Effect of hetroscedasticity on OLS results

لقد اوضحنا في الفصل الثاني بالنسبة للانحدار البسيط ان معلمة الانحدار:

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i u_i \quad , \qquad k_i = \left(\frac{x_i}{\sum x_i^2}\right)$$

وان تباين معلمة الانحدار :

$$\mathrm{var}\left(\hat{\beta}_1\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} = \sum k_i^2 \sigma_i^2$$

ففي حالة $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma^2$ يصبح التباين للمعلمة:

$$\mathrm{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum x_i^2} \qquad \cdots \qquad (4-6)$$

أما في حالة عدم التجانس فان التباين:

$$\mathrm{var}\left(\hat{\beta}_1\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \qquad \cdots \qquad (5-6)$$

ولابد من التأكيد على ان البرامجيات الجاهزة في حالة عدم التجانس ترتكب خطأين:

الأول: يتم باستخدام الصيغة (٦–٤) لحساب تباين معلمة الانحدار عوضاً عن الصيغة (5–6)

أما الأمر الثاني: فهو استخدام الصيغة $(\hat{\sigma})$ عوضاً عن $(\hat{\sigma}_i)$ ، حيث إن

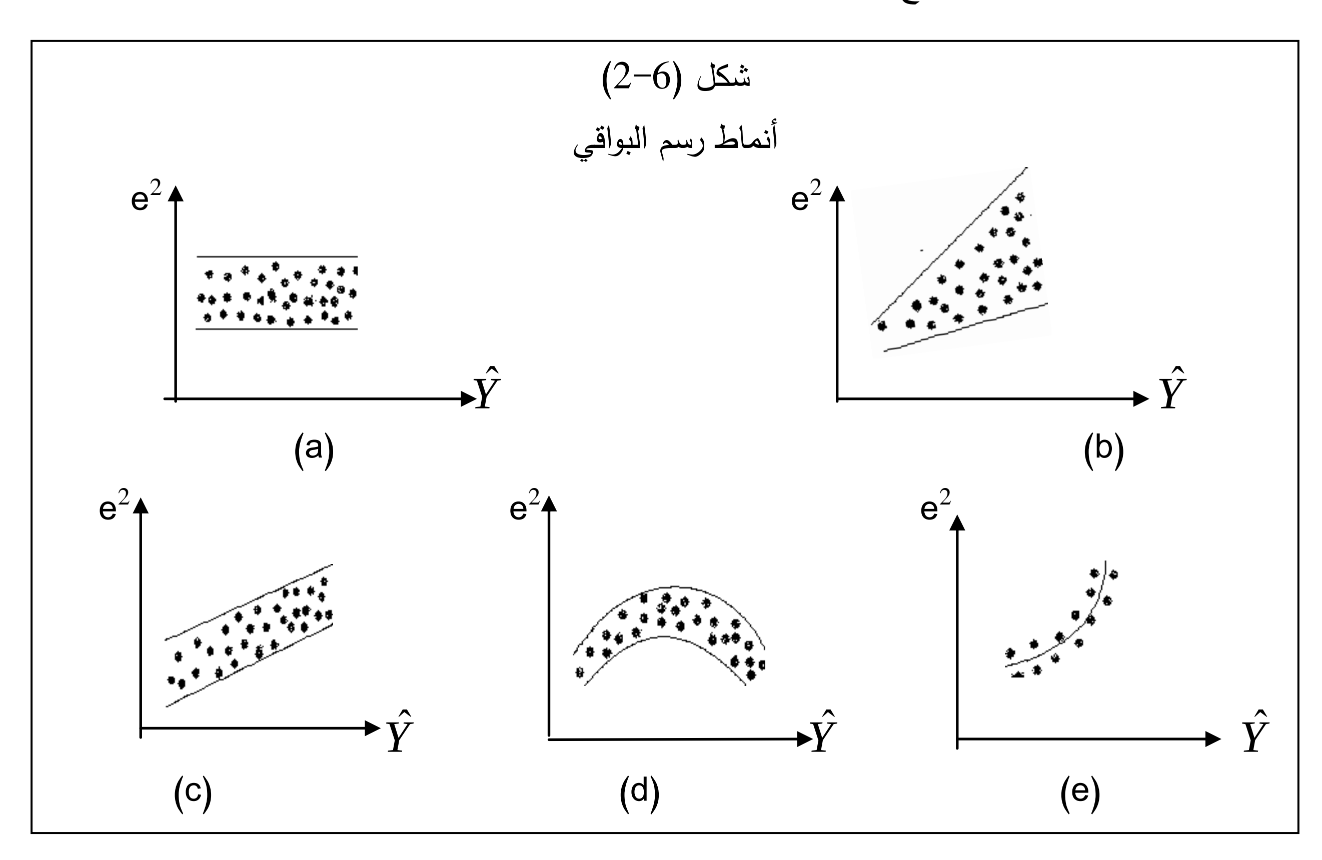
$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{\sum e_i^2}{n-2}$$

وبذلك فان تباين المعلمات يكون متحيزاً. في حين تبقى المقدرات تتصف بخاصية عدم التحيز وخطية العلاقة بدلالة Y كما إن المعلمات تكون متسقة غير أنها تفقد خاصية الكفاءة لان تباينها لا يكون اقل ما يمكن. وبذلك فان اختبارات المعنوية تكون متحيزة (لأنها تعتمد على تباين المعلمات) وكذلك مجالات الثقة للمعلمات المقدرة أو لمتوسط الاستجابة أو للقيمة التنبؤية الجديدة تكون متحيزة هي الأخرى ولا يعتد بها (تفقد مصداقيتها).

(٦-3-٤) طرائق الكشف عن المشكلة.Methods of Detections

في حال غياب أي معلومات حول طبيعة عدم التجانس، فان طريقة الرسم باستخدام بواقي علاقة الانحدار (والتي تعد تقديراً للمتغير العشوائي (u) وخاصة مع كبر حجم العينة) والتي تسمى تحليل البواقي (Residual Analysis) ،تعد مدخلا أوليا لتلمس ذلك. إذ يتم إجراء الانحدار بافتراض عدم وجود مشكلة عدم التجانس وتحسب البواقي $e_i = \hat{u}_i$ ، ثم يتم رسم (مربع البواقي) e_i^2 ، [يمكن أن يتم التحليل على أساس رسم قيم (e)] على المحور العمودي و (X) أو $\hat{Y}$ على المحور الأفقي.
فإذا كان الشكل يظهر نمطاً معيناً يمكن استنتاج وجود أو عدم وجود المشكلة وكذلك النمط المناسب لعدم التجانس. والأشكال التالية توضح ذلك:

شكل (6-2)
أنماط رسم البواقي



فالشكل (6-٢ a) يبين عدم وجود نمط محدد وذلك يقترح بان البيانات لا تظهر عدم تجانس لتغيرات الأخطاء. أما الأشكال المتبقية فتشير إلى وجود مشكلة عدم التجانس وبأنماط مختلفة:

الشكل (b) يبين إن تباين الأخطاء متزايد مع قيم Y المقدرة والشكل (c) يبدي نمطاً خطياً يبن مربع الأخطاء وقيم Y المقدرة. والشكل d ,e تؤكد وجود علاقة تربيعية بين تباين الأخطاء وبين قيم Y المقدرة.

وجدير بالذكر إن استخدام تحليل البواقي يساعد في تحديد نمط عدم التجانس وبذلك تكوين الخطوة الأولى في حل مشكلة عدم التجانس من خلال معرفة نمط عدم التجانس لتحويل المشاهدات. وتعد هذه الطريقة تأشيرية إذ إنها تعتمد على الحكم الشخصي وخاصة مع محدودية مشاهدات العينة. ويجب ان تقرن مع اختبارات إحصائية. والاختبارات الإحصائية متعددة منها ما يعتمد على التوزيع الطبيعي وبعضها عام ومن أهم الاختبارات الإحصائية :

١- اختبار بارك (1966)(Park–test):

افترض بارك ان (σ_i^2) دالة بدلالة X_i ، فقد افترض : $\sigma_i^2 = \sigma^2 X_i^{\beta} e^{u_i}$ وحيث ان (σ_i^2) غير معروفة بشكل عام، لذا فقد اقترح بارك استخدام بواقي علاقة الانحدار (e_i) عوضا عنها.

ويتم تقدير $\ln(e_i^2)$ بدلالة $\ln X_i$: $\ln e_i^2 = \ln \sigma^2 + \beta \ln X_i + u_i$

ويتم اختبار معنوية معلمة الانحدار، فإذا أثبتت معنويتها فذلك دليل على وجود مشكلة عدم التجانس.

٢- اختبار كليزر(1969)" Glejser "

اعتمد كليزر على المبدأ نفسه الذي اعتمده بارك. بعد الحصول على الخطأ (ei) يتم تقدير القيمة المطلقة للبواقي كتراكيب خطية بدلالة X ثم تختبر معنوية معلمة الانحدار باستخدام t . فمعنوية معلمة الانحدار تشير الى وجود مشكلة عدم التجانس وأهم الصيغ التي اعتمدها هي:

$$(a)\ |e_i| = \alpha_1 + \beta_1 X_i + v_i$$

$$(b)\ |e_i| = \alpha_2 + \beta_2 \sqrt{X_i} + v_2$$

$$(c)\ |e_i| = \alpha_3 + \beta_3 \frac{1}{X} + v_3$$

$$(d)\ |e_i| = \alpha_4 + \beta_4 \frac{1}{\sqrt{X}} + v_4$$

$$(e)\ |e_i| = \alpha_5 + \beta_5 X^2 + v_5$$

$$(f)\ |e_i| = \sqrt{\beta_1 + \beta_2 X} + v_6$$

$$(g)\ |e_i| = \sqrt{\beta_1 + \beta_2 X^2} + v_7$$

مع ملاحظة إن العلاقة (f),(g) غير خطية بدلالة المعلمات، لذا لا يمكن تقديرها بطريقة المربعات الصغرى الاعتيادية.

ويمكن اعتماد طريقة اختبار كليزر لمعرفة نمط عدم التجانس للمساعدة في تحويل المشاهدات.

مثال (3-6): من علاقة انحدار الإنفاق الاستهلاكي Y والإنتاجية X توفرت لوحة البيانات التالية:

$$|e_i| = 407.27 - 0.02X_i \qquad r^2 = 0.0127$$
$$s.e.: \ (633.2) \ (0.0675)$$
$$\ln(e_i^2) = 35.8 - 2.8\ln X_i$$
$$t \ : \ (0.934) \ (-0.667)$$

اختبر مشكلة عدم التجانس على وفق اختبار كليزر وكذلك اختبار بارك.

الحل:

على وفق اختبار كليزر: نحسب قيم t الحسابية لمعلمة الانحدار في العلاقة الأولى:

$$\left| t_{\hat{\beta}_1} \right| = \left| \frac{-0.02}{0.0675} \right| = 0.296$$

اذن معلمة الانحدار غير معنوية.

وذلك يعكس ان النموذج لا يعاني من مشكلة عدم التجانس، وكذلك على وفق طريقة بارك ، حيث ان قيمة t الحسابية لمعلمة الانحدار في العلاقة الثانية (-0.667) وهي اقل من قيمة t الجدولية بدرجات حرية (n−2) ومستوى دلالة ٠.٠٢٥، وعليه يمكن الاطمئنان بان النموذج المستخدم لا يعاني من مشكلة عدم التجانس.

٣-اختبار سبيرمان لارتباط الرتب Spearman's Rank correlation test

يعتمد هذا الاختبار على حساب علاقة الارتباط بين بواقي الانحدار وبين قيمة المتغير التوضيحي في الانحدار البسيط أو احد المتغيرات التوضيحية أو قيمة $\hat{Y}_i$ في الانحدار المتعدد.

على وفق القانون:

$$r_{eX} = 1 - 6\left[\frac{\sum d_i^2}{n(n^2 - 1)} \right] \qquad \cdots \qquad (6-6)$$

حيث إن:

d_i فرق الرتب ، n حجم العينة $(n > 8)$

ونلخص خطوات الاختبار:

- انحدار Y علىX وتستخرج البواقي (e_i).
- تحسب رتبة القيمة المطلقة للبواقي $(\rho|e_i|)$ وكذلك رتبة (X_i) $(\rho(X))$ او $(\hat{Y})$ $(\rho(\hat{Y}))$ في حالة الانحدار المتعدد ثم يطبق القانون (6-٦).
- نختبر الفرضية : $H_0 : \rho = 0 \quad vs \quad H_1 : \rho \neq 0$

باستخدام اختبار t يمكن معرفة معنوية r_{ex} (r_{ex} يشير الى معامل الارتباط البسيط بين البواقي (e) ومشاهدات المتغير X):

$$t^* = r_{eX}\sqrt{\frac{n-2}{1-r_{eX}^2}} \quad \sim \quad t_{c(n-2,\frac{\alpha}{2})}$$

معنوية r تدل على وجود مشكلة عدم تجانس تباين الخطأ وبعبارة أخرى:

اذا $t^* \rangle\ t_c \Leftarrow$ وجود مشكلة عدم التجانس.

ملاحظة: في حالة وجود قيم مكررة فان حساب الرتب يتم بجمع الرتب للقيم المتكررة ثم قسمتها على عدد التكرارات وتحدد الرتبة نفسها لكل قيمة من القيم المكررة.

مثال(6-4):الجدول في أدناه يوضح حساب معامل ارتباط الرتب لسبيرمان لعشر مشاهدات لقيم Y و X.

Y	12.4	14.4	14.6	16.0	11.3	10.0	16.2	10.4	13.1	11.3
X	12.1	21.4	18.7	21.7	12.5	10.4	20.8	10.2	16.0	12.0
$\|e\|$	1.03	1.24	0.20	0.22	0.26	0.59	0.83	0.10	0.06	`0.03
$\rho\|e\|$	9	10	4	5	6	7	8	3	2	1
$\rho(X)$	4	9	7	10	5	2	8	1	6	3
d	5	1	−3	−5	1	5	0	2	−4	−2
d^2	25	1	9	25	1	25	0	4	16	4

$\sum_i d^2 = 110$

$$r_{eX} = 1 - 6\frac{110}{10(100-1)} = 0.333$$

$$t^* = \frac{0.333\sqrt{8}}{\sqrt{1-0.111}} = 0.9998 \qquad , \ t_{c(8,0.225)} = 2.447$$

فيكون القرار: لا توجد مشكلة عدم تجانس التباين.

٤-أ ختبار جولدفيلد – كواندت test Goldfeld – Quandt

بافتراض ان تباين المتغير العشوائي (σ_i^2) مرتبط طردياً مع أحد المتغيرات التوضيحية

أي مثلاً : σ^2 ثابت ، $\sigma_i^{\,2} = \sigma^2 X_i^{\,2}$

خطوات الاختبار:

- ترتب المشاهدات على وفق اقل قيمة لـ X.
- تحذف (c) من المشاهدات الوسطية وتقسم المشاهدات إلى جزأين، ($\frac{n-c}{2}$) لكل منها. من أجل تكثيف الفرق بين المجموعتين،

ويتم اختيار c على وفق الاتي : $\begin{Bmatrix} c=4 \ \ if \ \ n=30 \\ c=8 \ \ if \ \ n=60 \end{Bmatrix}$

- يتم اجراء الانحدار لكل جزء من المشاهدات ،باستخدام OLS ، ويحسب مجموع مربعات الخطأ لكل جزء منها.

RSS_1: هي مجموع مربعات الاخطاء للمجموعة ذات التباين الاصغر.

RSS_2: هي مجموع مربعات الاخطاء للمجموعة ذات التباين الاكبر.

وبدرجات حرية متماثلة لكل جزء وهي:

$$\left(\frac{n-c}{2} - k\right) = \frac{n-c-2k}{2}$$

k: عدد المعلمات بضمنها المقطع الصادي.

- تحسب النسبة λ كالاتي:

$$\lambda = \frac{RSS_2 / d.f}{RSS_1 / d.f} \ \sim \ F_{c\left(\frac{n-c-2k}{2}, \frac{n-c-2k}{2} \ , \ 1-\alpha\right)}$$

ويكون القرار بان عدم التجانس موجود فيما اذا كان $\lambda > F_c$

ملاحظة: مدى نجاح الاختبار يعتمد على اختيار (c) وكذلك المتغير التوضيحي الذي يسبب المشكلة في حال وجود اكثر من متغير توضيحي تكرر العملية لكل متغير.

مثال(5−6): انحدار على أول (١٣) مشاهدة :

$$\hat{Y}_i = 3.4 + 0.69X_i \quad , \quad r^2 = 0.8887$$
$$s.e: \quad (0.07) \quad , \quad RSS_1 = 377.17 \quad , \quad d.f = 11$$

انحدار على آخر(١٣) مشاهدة :

$$\hat{Y} = -28.02 + 0.79X_i \quad , \quad r^2 = 0.768$$
$$s.e: \quad (0.13) \quad , \quad RSS_2 = 1536.8 \quad , \quad d.f = 11$$

لعينة حجمها 30 = n ، وعليه فان عدد المشاهدات الوسطية المحذوفة (c = 4)

وان النسبة λ المحسوبة هي:

$$\lambda^* = \frac{1536.8}{377.17} = 4.07 \quad , \quad F_{c(11,11,0.95)} = 2.82$$

وبذلك فان القرار يؤكد وجود مشكلة عدم تجانس تباين البواقي .

٥−اختبار بورش−بيجين −جود فري (Brush− Pagan − Godfrey) (BPG test(1979)

في النموذج الخطي المتعدد $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + u$

يفترض: $\sigma_i^2 = f(\alpha_0 + \alpha_1 z + \ldots + \alpha_k z_k)$

حيث ان z هي بعض او كل المتغيرات التوضيحيةX.

أي: $\sigma_i^2 = (\alpha_0 + \alpha_1 z + \ldots + \alpha_k z_k)$

اذا $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$ فان $(\sigma_i^2 = \alpha_0)$ أي ان التباين ثابت.

$$\Rightarrow \quad \sigma_i^2 = \alpha_0 \quad constant$$

خطوات الاختبار:

- نجري انحدار Y على كل $X\text{'}S$ ليتم الحصول على الأخطاء $e_1, \cdots, e_n$.
- يقدر $\hat{\sigma}_u^2 = \frac{\Sigma e_i^2}{n}$ والتي تمثل تقدير الإمكان الأعظم لتباين الخطأ.
- نعرف: $P_i = \frac{e_i^2}{\hat{\sigma}^2}$ ، أي قسمة e_i^2 على $\hat{\sigma}^2$ نرمز له P_i.
- نقدر P_i على: $Z\text{'}S$:

$$P_i = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \ . \ . \ . \ + \alpha_m z_m + v_i$$

- يتم الحصول على ESS (مجموع المربعات المشروحة) ومع افتراض تحقق التوزيع الطبيعي للمتغير u .

- تعرف الاحصاءة (BPG) كآلاتي: $BPG = \frac{1}{2} ESS$

- وتتوزع الاحصاءة (BPG) حسب توزيع مربع كآي $BPG \overset{asy}{\sim} \chi^2_{m-1}$

اذا $BPG > \chi^2_c \Leftarrow$ ***نفرض*** H_0 أي يكون القرار بوجود مشكلة عدم التجانس .

مثال(6−6):يتم استخدام لوحة البيانات التالية لتوضيح فكرة اختبار بورش−جود فري−بيجين :

- انحدار Y على X يوفر المعلومات التالية:

$$\hat{Y}_i = 9.29 + 0.64X_i \quad , \quad i = 1,...,30 \quad , \quad RSS = 2361.2$$

$$s.e \quad (0.028) \quad , \quad R^2 = 0.9466$$

- تحسب (p_i) وعلى وفق الآتي :

$$\hat{\sigma}^2 = \frac{RSS}{n} = \frac{2361.2}{30} = 78.7$$

$$P_i = \frac{e_i^2}{\hat{\sigma}^2} = \frac{e_i^2}{78.7}$$

- نفترض P_i تركيب خطي بدلالة ($z_i = X_i$)

- ونجري انحدار P_i على X_iفنحصل على المعلومات:

$$\hat{P}_i = -0.74 + 0.01X_i$$

$$s.e_{-} \quad (0.004)$$

$$ESS = 10.428 \quad , \quad R^2 = 0.18$$

- نحسب قيمة الأحصاءة ونقارنها مع مربع كاي بمستوى دلالة 5٪ وبدرجة حرية (١):

$$\Rightarrow BPG = \frac{1}{2} ESS = 5.214 \overset{asy}{\sim} \chi_1^2 = 3.841$$

- وحيث أن : $BPG > \chi 2$

- فيكون القرار: برفض H_0 أي توجد مشكلة عدم التجانس.

ملاحظة: لابد من ملاحظة إن الاختبار يتحقق بالنسبة للعينات الكبيرة إذ يتطلب الاختبار افتراض التوزيع الطبيعي.

٦- اختبار وايت لعدم التجانس العام. (١٩٨٠) White test

مما تقدم يتضح ان اختبار كولد فيلد - كواندت يتطلب معرفة أي متغير توضيحي هو الذي يسبب المشكلة. واختبار(BPG) يفترض التوزيع الطبيعي للمتغير العشوائي ، او بافتراض كبر العينة.

في حين نجد ان اختبار white عام . وبافتراض النموذج:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

فان خطوات الاختبار

١- نقدر النموذج للحصول على البواقيe_i.

٢- نجري انحدار $(e_i)^2$ على $X_1 X_2, X_2^2, X_1^2, X_2, X_1$ والثابت

$$e_i^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1 X_2 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + v_i$$

٣- نحسب قيمة R^2 من معادلة الانحدار، ثم (nR^2) .

٤- نحسب الاحصاءة ، وتتوزع الاحصاءة حسب توزيع مربع كاي: $nR^2 \overset{asy}{\sim} \chi_5^2$

وبشكل عام فان $nR^2 \sim \chi^2_{\frac{k(k+3)}{2}}$، حيث k تمثل عدد المتغيرات التوضيحية في النظام .

٥- القرار: اذا $nR^2 > \chi_c^2 \Leftarrow$ توجد مشكلة عدم التجانس

علما ان الفرضية هي:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_5 = 0$$ التباين متجانس

$$H_1 : \alpha_1 \neq \alpha_2 \neq ... \neq \alpha_5 \neq 0$$ التباين غير متجانس

مثال(6-7): - من بواقي العلاقة:

$$Y = \beta_1 + \beta_2 \ln X_1 + \beta_2 \ln X_1 + \beta_2 \ln X_2 + u_i \quad ,n = 41$$

- تم تقدير مربع البواقي على كل من المتغيرات:

$$(\ln X_1)((\ln X_2), \quad (\ln X_2)^2, (\ln X_1)^2, \quad \ln X_1, \ln X_2$$

- فتم الحصول على النتائج:

$$\hat{e}_i^2 = -5.8 + 2.56 \ln X_1 + 0.69 \ln X_2 \qquad ,$$
$$-0.4 \ (\ln X_1)^2 - 0.04 \ (\ln X_2)^2$$
$$+0.001 \ (\ln X_1)(\ln X_2) \quad ; \quad R^2 = 0.1148$$

- وبذلك فان الاحصاءة white :

$$\Rightarrow nR^2 = (41)(0.1148) = 4.7068$$

وبمقارنتها مع مربع كاي الجدولية بمستوى دلالة ٥% و5 درجات حرية $\chi_5^2 = 11.0705$

فيكون القرار بموجب اختبار White لا توجد مشكلة عدم التجانس.

مثال (6−8) استخدم بيانات المثال (3−4) نفسها لاختبار مشكلة عدم التجانس باعتماد اختبار white وذلك بالاعتماد على برنامج SPSS.

الحل: بعد حساب القيمة المقدرة لـ $\hat{Y}$ منها تحسب e_i حيث ان $e_i = Y_i - \hat{Y}_i$

− يحسب e_i^2 و X^2

− ثم تقدر e_i^2 بدلالة المقطع الصادي و X و X^2 .

أي: $e_i^2 = \beta_0 + \beta_1 X + \beta_2 X^2 + u$ وتحسب الاحصاءة وايت: $\{(W) = nR^2\}$

ونتائج SPSS تعرض في الجدول (6−5)

جدول (6−5)

نتائج التحليل الخاصة بالمعلمات المقدرة

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	β	Std. Error	Beta		
(Constant)	−13.848	18.803		−.737	.476
X	5.187	5.052	2.243	1.027	.325
X^2	−.287	.302	−2.081	−.953	.359

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

معادلة التقدير:

$$e_i^2 = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2$$

$$= -13.848 + 5.187\, X_2 - 0.287\, X^2$$

غير ان الذي يهمنا لحساب اختبار وايت هو معامل التحديد للنموذج المذكور آنفاً والذي يمكن استخراجه من الجدول (6−6)

جدول (6-6)

ملخص النموذج

R	R Square (R^2)	Adjusted R Square ($\bar{R}^2$)	Std. Error of the Estimate	Durbin-Watson
.316	.100	-.050	-.050	1.885

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

وحيث ان حجم العينة n = 15W = n R^2

= (15) (0.100) = 1.5

وتقارن قيمة الاحصاءة مع قيم مربع كاي وبدرجات حرية 2 ومستوى معنوية 5% وهو:

$$\chi^2_{c(2,0.05)} = 5.99$$

نلاحظ ان قيمة W الحسابية اقل من χ^2 الجدولية لذا نقبل فرضية العدم وهذا يعني انه لاتوجد مشكلة عدم التجانس في تباين بواقي النموذج.

(6-3-٥) طرائق علاج مشكلة عدم التجانس.Remedial measures

بالرغم من ان مشكلة عدم التجانس لا تلغي خاصية عدم التحيز والاتساق للمعلمات المقدرة غير انها (المعلمات المقدرة) تصبح غير كفوءة، وان تباين المعلمات المقدرة يكون متحيزا وبدوره يؤثر في قيم t لهذه المعلمات المقدرة وتصبح محل شك. لذا لابد من أيجاد طريقة للمعالجة. ومحور طرق المعالجة هي تحويل جميع مشاهدات المتغيرات المستخدمة (بضمنها المقطع الصادي) بالشكل الذي يضمن تخليص المتغير العشوائي من مشكلة عدم التجانس، ثم يقدر النموذج المحول بطريقة المربعات الصغرى الاعتيادية وتسمى هذه الطريقة طريقة المربعات الصغرى الموزونة (المرجحة).

سيتم توضيح الفكرة بالنسبة للنموذج البسيط:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad , \quad \operatorname{var}(u_i) = \sigma_i^2$$

أولاً: عندما σ_i^2 تكون معلومة فان التحويل يتم بقسمة مشاهدات كل متغير على الانحراف المعياري للخطأ (σ_i)، وبذلك فان النموذج المحول:

$$\frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

وبافتراض ان

$$\frac{1}{\sigma_i} = z_i^* \ , \ \frac{Y_i}{\sigma_i} = Y_i^* \ , \ \frac{X_i}{\sigma_i} = X_i^*, \ \frac{u_i}{\sigma_i} = u_i^*$$

فالنموذج المحول :

$$Y_i^* = \beta_1^* z_i^* + \beta_2^* X_i^* + u_i^*$$

$$\text{var}(u_i^*) = \text{var}\left(\frac{u_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2}\text{var}(u_i) = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

أي ان تباين المتغير العشوائي في النموذج المحول يكون متجانساً. لذا يتم تقدير معلمات النموذج المحول بواسطة طريقة المربعات الصغرى الاعتيادية للحصول على $\hat{\beta}_2^*, \hat{\beta}_1^*$ والتي تكون BLUE. ثم ربطها بقيم β_0 و β_1.

ثانيا: عندما تكون σ_i^2 غير معروفة فيمكن إتباع الخطوات التالية لتقديرها.:

١- طريقة White للعينات الكبيرة. وتذكر في العديد من البرامجيات الجاهزة وتتلخص كآلاتي:

اقترح White تقدير لـ σ_i^2 باستخدام مربع البواقي المقدرة $(\hat{e}_i^2)$

حيث ان: e_i هي بواقي علاقة انحدار Y على جميع المتغيرات التوضيحية($X\,'S$) أما معلمات الانحدار فتحسب تبايناتها في حالة الانحدار البسيط على وفق:

$$\text{var}(\hat{\beta}_1) = \frac{\sum x_i^2 e_i^2}{\left(\sum x_i^2\right)^2}$$

اما في حالة وجود k من المتغيرات التوضيحية فان

$$\text{var}(\hat{\beta}_j) = \frac{\sum \hat{w}_j^2 e_i^2}{\left(\sum \hat{w}_j^2\right)^2}$$

(e_i^2) هي بواقي انحدار Y على جميع المتغيرات التوضيحية ($X_1, \ldots, X_k$)

$\hat{w}_j$ هي بواقي انحدار X_j على باقي المتغيرات التوضيحية.

$$X_j = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \ldots + \alpha_k X_k + w$$

١. يتم تقدير النموذج للحصول على البواقي (e_i) ثم نقدر مربع البواقي (e_i^2) كتركيب لمتعدد حدود من الدرجة k ونستخرج بواقي العلاقة $(\hat{v}_i)$.

$$e_i^2 = a_0 + a_1 X + a_2 X^2 + \ldots a_k X^k + v_i$$

ويتم استخدام δ_i^2 كتقدير σ_i^2 ، حيث ان

$$\hat{\delta}_i^2 = \frac{e_i^2}{\hat{v}_i^2}$$

٢- معرفة نمط عدم التجانس من خلال اختبار Park او Glejser او من خلال تحليل البواقي وكما في الشكل (2-6).

٣- في حالة مشاهدات Y_i هي عبارة عن متوسطات فان σ_i^2 تمثل تباين الوسط الحسابي $\sigma_i^2 = \frac{\hat{\sigma}^2}{n_p}$.

أي ان

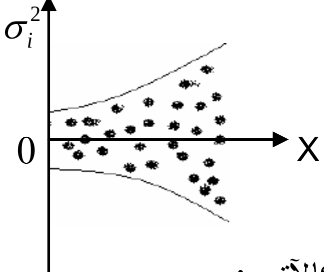
$$\Omega = \begin{pmatrix} \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & 0 & \frac{1}{n_p} \end{pmatrix}$$

حيث ان p: تمثل عدد المجموعات.

و n_p: عدد مشاهدات المجموعة p

وبعد تقدير قيمة σ_i^2 يتم تحويل المشاهدات كما في (أولاً) والأمثلة توضح ذلك .

مثال (6-9): عند معرفة نمط التجانس $E(u_i)^2 = \sigma^2 X_i^2$ بالاعتماد على رسم مربع البواقي ضد X يتضح من الشكل المجاور ان (التباين يتناسب مع مربع X)



لذا فان التحويل المناسب يكون بقسمة الطرفين على قيم X_i كالآتي :

$$\frac{Y_i}{X_i} = \frac{\beta_0}{X_i} + \beta_1 + \frac{u_i}{X_i}$$

$$Y^* = \beta_0 X^* + \beta_1 + u^*$$

اذ يكون $\mathrm{E}(u^*) = \sigma^2$ متجانساً وبذلك تكون المعادلات الطبيعية:

$$\Sigma Y^* = n\hat{\beta}_1 + \hat{\beta}_0 \Sigma X^*$$

$$\Sigma X^* Y^* = \hat{\beta}_1 \Sigma X^* + \hat{\beta}_0 \Sigma X^{*2}$$

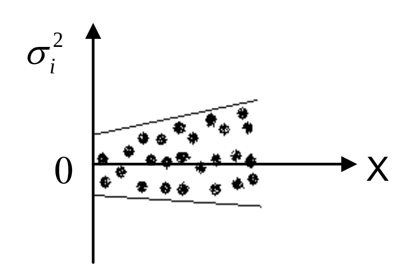
أي أن المعادلات الطبيعية:

$$\Sigma \frac{Y_i}{X_i} = n\hat{\beta}_1 + \hat{\beta}_0 \Sigma \frac{1}{X_i}$$

$$\Sigma \frac{Y_i}{X_i} \cdot \frac{1}{X_i} = \hat{\beta}_1 \Sigma \frac{1}{X_i} + \hat{\beta}_0 \Sigma \frac{1}{X_i^2}$$

وبحل المعادلات نحصل على المعلمات المقدرة للنموذج الأصلي $\hat{\beta}_1, \hat{\beta}_0$.

مثال (6−10): نفترض (التباين يتناسب مع X) كما يوضحه رسم تحليل البواقي المجاور:



وبذلك : $E(u_i^2) = \sigma^2 X_i$

فان التحويل المناسب يكون بقسمة الطرفين على $(\sqrt{X_i})$:

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_0}{\sqrt{X_i}} + \beta_1 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}}$$

حيث:

$$u_i^* = \frac{u_i}{\sqrt{X_i}} \text{ ، } X_{2i}^* = \sqrt{X_i} \text{ ، } X_{1i}^* = \frac{1}{\sqrt{X_i}} \text{ ، } Y_i^* = \frac{Y_i}{\sqrt{X_i}}$$

اي:

$$Y_i^* = \beta_0 X_{1i}^* + \beta_1 X_{2i}^* + u_i^*$$

اي ان علاقة الانحدار بدون مقطع صادي وبذلك

$$e^* = (Y^* - \hat{\beta}_0 X_1^* + \hat{\beta}_1 X_2^*)$$

$$Q = \Sigma e^{*2} = \Sigma (Y^* - \hat{\beta}_0 X_1^* + \hat{\beta}_1 X_2^*)^2$$

$$\frac{\partial Q}{\partial \beta_0} = 2\Sigma(Y^* - \hat{\beta}_0 X_1^* - \hat{\beta}_1 X_2^*) \cdot (-X_1^*) = -2\Sigma(X_1^* Y^* - \hat{\beta}_0 X_1^{*2} - \hat{\beta}_1 X_1^* X_2^*) = 0$$

$$\frac{\partial Q}{\partial \beta_1} = 2\Sigma(Y^* - \hat{\beta}_0 X_1^* - \hat{\beta}_1 X_2^*) \cdot (-X_2^*) = -2\Sigma(X_2^* Y^* - \hat{\beta}_0 X_1^* X_2^* - \hat{\beta}_1 X_2^{*2}) = 0$$

اذن المعادلات الطبيعية:

$$\left.\begin{aligned} \Sigma X_1^* Y^* &= \hat{\beta}_0 \Sigma X_1^{*2} + \hat{\beta}_1 \Sigma X_1^* X_2^* \\ \Sigma X_2^* Y^* &= \hat{\beta}_0 \Sigma X_1^* X_2^* + \hat{\beta}_1 \Sigma X_2^{*2} \end{aligned}\right\} \quad \ldots \quad (7-6)$$

لجميع قيم i و j ، i , j =1, . . . , n

وبصيغة المصفوفات:

$$\begin{bmatrix} \Sigma X_1^* Y^* \\ \Sigma X_2^* Y^* \end{bmatrix} = \begin{bmatrix} \Sigma X_1^{*2} & \Sigma X_1^* X_2^* \\ \Sigma X_1^* X_2^* & \Sigma X_2^{*^2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$X^{*'} Y^* = (X^{*'} X^*)^{-1} \hat{\beta}$$

مثال(6 −11): بافتراض: $\mathrm{E}(u_i^2) = \sigma^2 \mathrm{E}(Y_i)^2$

اي تباين المتغير العشوائي يتناسب مع مربع (توقع Y)

حيث ان توقع Y هو: $E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X$

اذن النموذج المحول:

$$\frac{Y_i}{E(Y_i)} = \beta_0 \frac{1}{E(Y_i)} + \beta_1 \frac{X}{E(Y_i)} + \frac{u_i}{E(Y_i)}$$

وحيث ان : $E(Y_i) = \hat{Y}_i$

اذن النموذج المحول:

$$\frac{Y_i}{\hat{Y}_i} = \beta_0 \frac{1}{\hat{Y}_i} + \beta_1 \frac{X_i}{\hat{Y}_i} + \frac{u_i}{\hat{Y}_i}$$

حيث ان $E\left(\frac{u_i}{E(Y_i)}\right)^2 = \sigma^2$ يكون متجانساً، لذا فالتحويل عمل على تخليص النموذج من مشكلة عدم التجانس.

$$Y_i^* = \beta_0 X_1^* + \beta_1 X_2^* + u_i^*$$

حيث ان:

$$Y_i^* = \frac{Y_i}{\hat{Y}_i} \quad ، \quad X_1^* = \frac{1}{\hat{Y}_i} \quad ، \quad X_2^* = \frac{X_i}{\hat{Y}_i} \quad ، \quad u^* = \frac{u_i}{\hat{Y}_i}$$

فيتضح نموذج الانحدار بدلالة متغيرين X_1^* و X_2^* وبدون مقطع صادي وبذلك فان المعادلات الطبيعية كما في (6−7).

ويلاحظ ان التحويل يعمل على تغيرات تؤثر في تفسير المعلمات ففي المثال (6−9) يكون المقطع الصادي للنموذج المحول هو β_1 في حين معلمة الانحدار هي β_0 .

وكذلك في المثالين (6−10) و (6−11) فان معلمة الانحدار للمتغير التوضيحي الأول هي β_0 في حين معلمة الانحدار للمتغير التوضيحي الثاني هي β_1 .

لذا بعد التقدير للنموذج المحول يعاد تعويض قيم المتغيرات كل حسب الافتراض المسبق لها.

- كما إن هناك طرائق علاج أخرى تؤكد محاولة ترميم الخلل الذي أدى إلى ظهور مشكلة عدم التجانس وذلك من خلال إتباع الآتي:

أ- في حالة كون مشكلة عدم التجانس سببها حذف متغيرات مهمة من النموذج فتتم المبادرة لإضافة متغيرات جديدة للنموذج.

ب - اذا كانت الصيغة الخطية هي المستخدمة فيتم استخدام الصيغة اللوغاريتمية المزدوجة ثم نختبر البواقي الناتجة فيما إذا كانت خالية من عدم التجانس. أو تحويل بعض المتغيرات في النموذج إلى صيغ أخرى. غير إن التحويل اللوغاريتمي لا يمكن تطبيقه إذا كانت احد قيم المتغيرات Y أو X تساوي صفرا أو قيمة سالبة.

هذا فضلا عن إن طريقة تحويل المشاهدات لتخليص النموذج من مشكلة عدم التجانس لا تخلو من مشكلات ومنها حدوث مشكلة الارتباط الزائف

فمثلا عند تحويل النموذج : $Y = \beta_0 + \beta_1 X + u$

كالآتي:

$$\frac{Y}{X} = \beta_0 \frac{1}{X} + \beta_1 + v$$

فقد تكون Y غير مرتبطة بـ X ولكن $\frac{Y}{X}$ مرتبطة بـ $\frac{1}{X}$

ج- طريقة المربعات الصغرى المعممة (GLS) وسيتم عرضها بشيء من التفصيل في المبحث(6-5).

(6-4)الارتباط الذاتي (Autocorrelation)

من الضروري توضيح بعض المفاهيم في هذا الصدد.

لقد ميز Titner (1965) ، بين مصطلحين هما: "Autocorrelation" و Serial Correlation .

فالارتباط الذاتي (Autocorrelation) هو ارتباط متباطئ لسلسلة معينة مع ذاتها وذلك من خلال تباطئها بعدد من الابطاءات الزمنية. في حين الارتباط المتسلسل هو ارتباط متباطئ بين سلسلتين مختلفتين، ولتوضيح ذلك نفترض سلسلتان زمنيتين مختلفتين بعشرة مشاهدات هما u و v

$$v_t = \{ v_1, v_2, \ldots, v_{10} \} \quad , \quad t = 1,2,\ldots,10$$

$$u_t = \{ u_1, u_2, \ldots, u_{10} \} \quad , \quad t = 1,2,\ldots,10$$

وبتباطؤ كل منهما بفترة إبطاء واحدة تتولد السلسلتين u_{t-1}و v_{t-1} :

$$u_{t-1} = \{ u_2, u_3, \ldots, u_{11} \}$$

$$v_{t-1} = \{ v_2, v_3, \ldots, v_{11} \}$$

فان ارتباط السلسةu_t مع u_{t-1} يسمى ارتباط ذاتي (Autcorrelation) في حين ارتباط السلسة u_t مع vt-1 يسمى ارتباط متسلسل (Serial autcorrelation).

في هذا المبحث سيتم التركيز على الارتباط الذاتي من خلال الإجابة عن الأسئلة التالية :

١. ما طبيعة الارتباط الذاتي؟

٢. ما الآثار النظرية و العملية للارتباط الذاتي؟

٣. ان فرضية الارتباط الذاتي تخص المتغير العشوائي، و هو متغير غير مشاهد، فكيف يمكن الاستدلال حول تحقق الفرضية من عدمها؟

٤. كيف نضع حلولاً لمعالجة مشكلة الارتباط الذاتي؟

(6-4-١) طبيعة المشكلة Natur of the problem

مصطلح الارتباط الذاتي :" Autocorrelation " يتم تعريفه بأنه ارتباط بين عناصر سلسلة من مشاهدات مرتبة عبر الزمن (في بيانات السلاسل الزمنية) أو مرتبة عبر المكان (في بيانات المقاطع العرضية). و في سياق نموذج الانحدار الخطي التقليدي فان الفرض (٥) ينص على إن لا وجود لترابط ذاتي بين مشاهدات المتغير العشوائي و بالرموز:

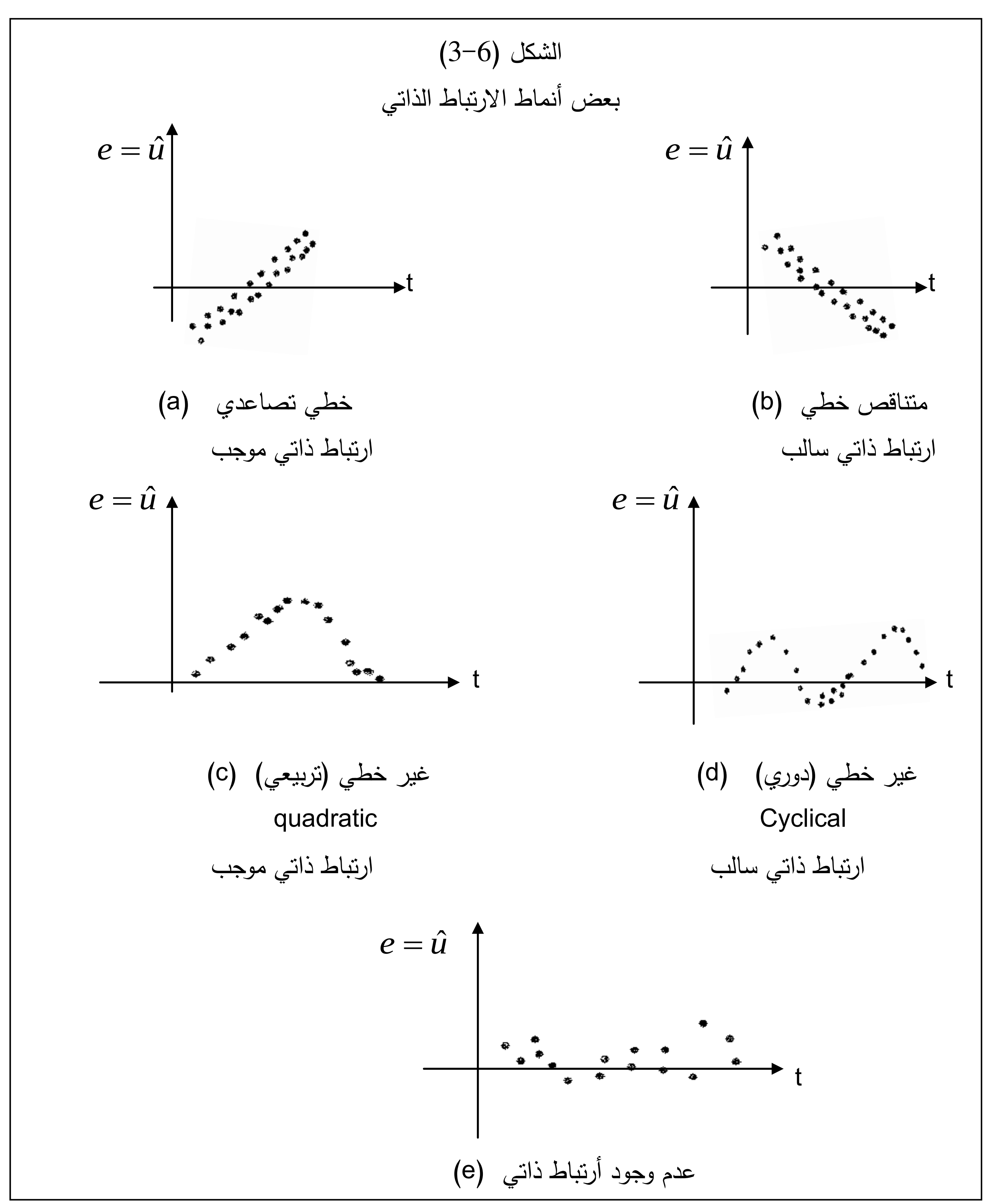
$$\mathrm{E}(u_i u_j) = 0 \qquad \mathrm{i} \neq \mathrm{j}, i, j = 1,2,....., n............(8-6)$$

و لتوضيح طبيعة المشكلة في بيانات سنوية نفترض علاقة انحدار بين ناتج عملية إنتاجية و بين مدخلات هذه العملية وهما العمل و رأس المال. و نفترض استخدام بيانات فصلية (ربع سنوية). نفترض وجود إضراب للعمال أثر في الناتج في فصل معين من الفصول. في هذه الحال، لا يوجد سبب للاعتقاد بان ذلك الأثر سيحمل آثاره فى الفصل اللاحق. فإذا تسبب إضراب العمال في فصل معين إلى خفض الناتج، فلا يوجد سبب بان نتوقع إن الناتج ينخفض في الفصل اللاحق. أما في حال وجود ارتباط ذاتي، بمعنى إذا العلاقة(٦-٨) لا تتحقق، فان أثر الإضراب العمالي لهذا الفصل ربما و بشكل أكيد يستمر تأثيره فى الناتج في الفصل اللاحق.

وبالمثل في دراسات المقاطع العرضية، لنفترض علاقة انحدار الإنفاق الاستهلاكي للعوائل بدلالة الدخل العائلي، فان اثر زيادة دخل إحدى العوائل على إنفاقها الاستهلاكي لا يتوقع إن يجري تأثيره فى الإنفاق الاستهلاكي للعوائل الأخرى. بينما في حال عدم تحقق العلاقة (٦-٨) فان زيادة الإنفاق الاستهلاكي لعائلة في العينة ربما تحفز عائلات أخرى لزيادة استهلاكهم من باب التقليد و المباهاة (keep up with the Joneses).

(٦-٤-٢) أنماط الارتباط الذاتي.Pattern of autocorrelation

هناك العديد من أنماط الارتباط الذاتي منها الخطية وغير الخطية. والخطية قد تكون تصاعدية أو اتجاه زمني خطي متناقص. كما إن الأنماط غير الخطية قد تكون تربيعيه أو دورية ويمكن عرض هذه الأنماط من خلال رسم بواقي علاقة الانحدار (e_t) ضد الزمن (t) و كما في الشكل (٦-3)



فالاشكال: (a) تشير الى وجود ارتباط ذاتي خطي متصاعد موجب.

(b) توضح ان هناك ارتباطاً ذاتياً خطياً متناقصاً سالباً.

(c) و (d) كل منهما تلمح الى وجود ارتباط ذاتي غير خطي تربيعي ودوري على التوالي.

فيما يشير الشكل في (e) الى عدم وجود ارتباط ذاتي.

(3-4-6)أسباب ظهور الارتباط الذاتي Causes of autocorrelation

1- القصور. "Inertia"

اغلب المتغيرات الاقتصادية مثل GNP ، الأسعار القياسية ، الإنتاج ، البطالة ، والتشغيل تظهر دورات. تبدأ من أسفل الركود ثم بالانتعاش اغلب هذه الدورات تبدأ بالصعود وفي هذه الحركة الصعودية فان قيم السلسلة في نقطة معينة من الزمن تكون اكبر من قيمتها السابقة. وبذلك يتولد عزم في هذه المتغيرات و يستمر ذلك إلى إن يحصل شيء ما (مثلاً زيادة سعر الفائدة أو الأسعار أو كلاهما) فتبدآ هذه المتغيرات بالانخفاض و عليه في حالة السلاسل الزمنية، فان المشاهدات المتتالية تكون في الغالب مترابطة فيما بينها .

2- تحييز التوصيف: ويظهر سوء التوصيف في حالتين:

(أ) - في حالة حذف متغيرات من التوصيف .

في الإشكال السابقة (d – a) تقترح بان بعض المتغيرات تم حذفها من النموذج.

فإذا كان النموذج الأصلي هو:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u \quad . \; . \; . \qquad (9-6)$$

و لسبب ما تم استخدام النموذج :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + v \quad . \; . \; . \qquad (10-6)$$

فإذا كانت العلاقة (6-9) هي النموذج الصحيح أو العلاقة الحقيقية فان استخدام العلاقة (6-10) تعني ضمنياً افتراض $v = \beta_3 X_3 + u$. وبذلك فان المتغير العشوائي v سوف يعكس نمطاً منتظماً مما يولد ارتباطاً ذاتياً كاذباً.

(ب) بسبب استخدام صيغ دالية غير مناسبة .

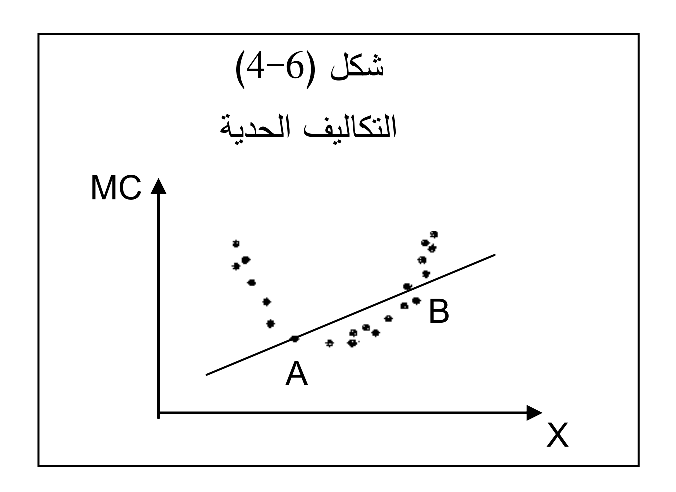
إذا كان النموذج الحقيقي للتكاليف الحدية

$$MC = \beta_1 + \beta_2 X + \beta_3 X^2 + u \quad . \; . \; . \qquad (11-6)$$

و لكن تم استخدام

$$MC = \alpha_1 + \alpha_2 X + v \qquad . \; . \; . \qquad (12-6)$$

فان ذلك يولد تحيزاً حيث إن ($v = \beta_3 X^2 + u$) يعكس نمطاً منتظماً يرتبط بمربع X وهو (الناتج) مما يولد ارتباطاً ذاتياً، بسبب استخدام التوصيف الدالي الخاطئ .



من خلال الشكل (6-4) فان القيم ما بين النقطتين A و B تعكس ان MC الخطية ستكون متحيزة للأعلى عن التكلفة الحدية الحقيقية، و فيما عدا هذه المسافة فإنها تكون متحيزة للأسفل عن القيمة الحدية الحقيقية.

٣- ظاهرة النسيج العنكبوتي (Cobweb Model): عرض اغلب السلع الزراعية يعكس ظاهرة ما يسمى النسيج العنكبوتي لان العرض يتأثر بالأسعار لسنة متباطئة لان قرارات العرض تتطلب زمناً لتحقيقها، (فترة التفريخ). ففي بداية سنة الزراعة، يكون المزارع متأثراً بالأسعار السائدة في السنة السابقة لذلك فان دالة العرض الزراعي :

$$Y_t = \beta_1 + \beta_2 P_{t-1} + u_t \quad . \; . \; . \quad (13-6)$$

وبافتراض إن السعر (P_t) اقل من (P_{t-1}) ، لذا في السنة (t + 1) يقرر المزارع التقليل من الزراعة عما كان علية في السنة (t) وبذلك فان(u_t) لا يتوقع إن يكون عشوائياً لان إذا زاد المزارع إنتاجه في السنة t فأنة سوف يقلص أنتاجه في السنة (t+1) وهكذا على وفق مبدأ النسيج ألعنكبوتي.

٤- الإبطاء (Lags)

وجود متغير داخلي متباطئ مثلاً في دالة الاستهلاك، تكون دالة بدلالة الدخل و الاستهلاك السابق للدلالة على العادات و التطوير التقني ولأسباب مؤسساتية وفسيولوجية.

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t \quad . \; . \; . \quad (14-6)$$

وهذا النموذج يسمى الانحدار الذاتي (auto regression)

فعند حذف هذا المتغير العشوائي (Y_{t-1}) فان المتغير العشوائي سوف يعكس نمطاً منتظماً بسبب أثر الاستهلاك السابق في الاستهلاك الحالي.

$$Y_t = \beta_1 + \beta_2 X_t + v_t \quad . \; . \; . \quad (15-6)$$

$$v_t = \beta_3 Y_{t-1} + u_t$$

٥- التعديلات في البيانات.

في الدراسات التطبيقية غالباً ما يلجأ إلى تعديلات في البيانات الأولية. مثلاً لدراسة انحدار السلاسل الزمنية ربع السنوية يتم إضافة بيانات ثلاثة أشهر وقسمة المجموع على (3)، وإن عملية استخدام المتوسطات سيولد تنعيم للبيانات. وبذلك فان الرسم للبيانات الفصلية تظهر تنعيماً أكثر من المشاهدات الشهرية. وهذا التنعيم نفسه سيمنح نمطاً متماثلاً للبواقي و بذلك تتسبب بالارتباط الذاتي.

كما إن نوع آخر من التعديل للمشاهدات هو (interpolation) أو (extrapolation)، مثلاً إذا كانت البيانات عن التعداد السكاني تعد كل (١٠) سنوات في العراق و آخر تعداد كان عام ٢٠٠٠ و التعداد السابق كان ١٩٩٠، فإذا هناك حاجة للحصول على معلومات للفترة(١٩٩٠-٢٠٠٠) فان الطريقة المعتادة هي interpolation على أساس فرضيات معينة، مثلاً باستخدام خط الاتجاه العام الخطي أو ألتربيعي أو الدالة اللوجستية، التي ستدخل على البيانات نمطاً معيناً ربما لا يكون موجوداً في البيانات الأصلية.

٦- تحويل المشاهدات (Transformation)

تحويل المشاهدات قد تكون مصدراً خصباً للارتباط الذاتي .

مثلاً: $Y_t = \beta_1 + \beta_2 X_t + u_t$

u_t غير مرتبط ذاتياً.

في حين يفضل في بعض الدراسات التطبيقية استخدام الصيغة:

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-2} + u_{t-1} \quad \cdots \quad (16 \quad -6)$$

أو بصيغة الفروق (الفرق الأول)

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t \quad \cdots \quad (17-6)$$

$$\Delta\, u_t = v_t$$

فإذا كانت Y: الإنفاق الاستهلاكي
و X: الدخل
بصيغة اللوغاريتمات

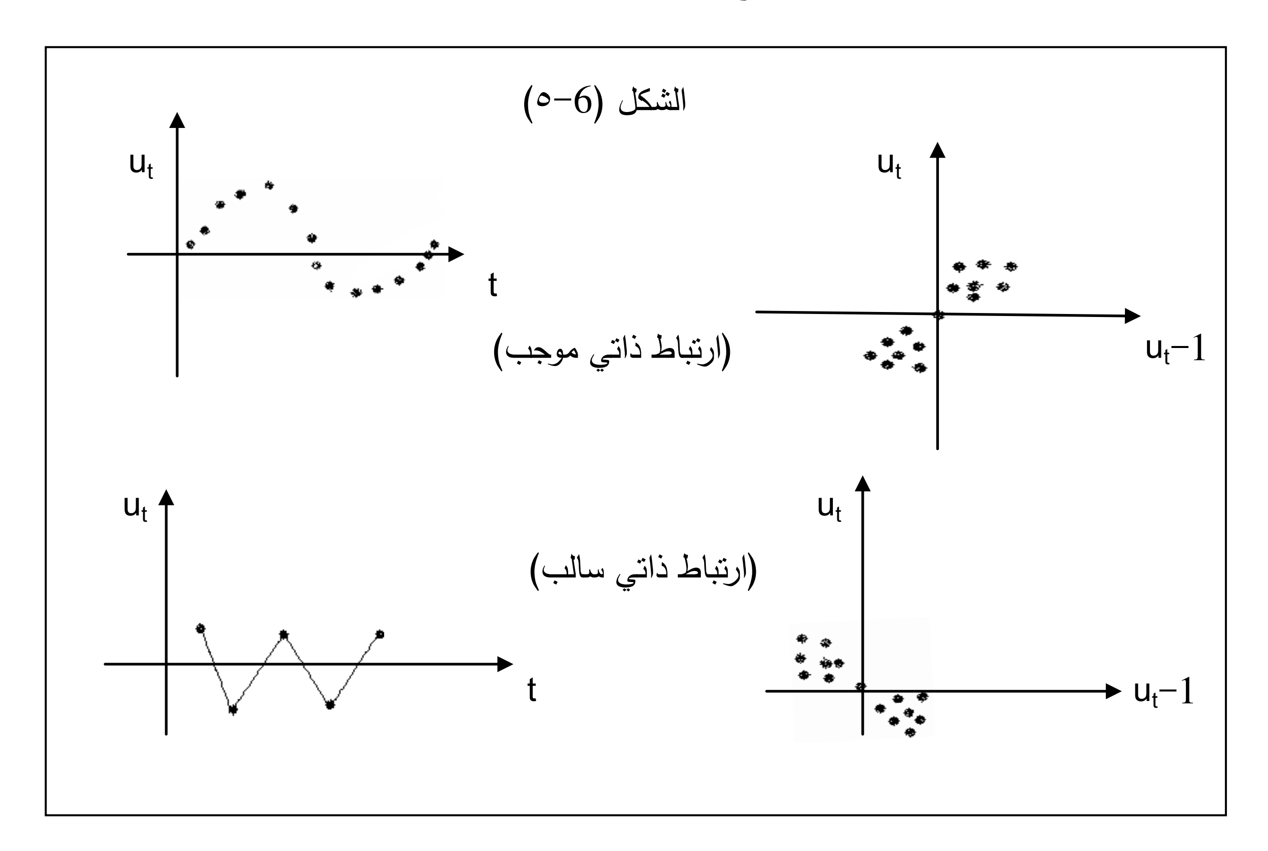
فان(٦-١٧) تمثل التغيرات في لوغاريتم الاستهلاك والدخل.

وحيث ان التغير في لوغاريتم المتغير تمثل التغير النسبي إذا تم ضربه × ١٠٠.

لذلك عوضاً من دراسة العلاقة بين المتغيرات بصيغة المستوى (level) ربما نكون مهتمين بدراسة العلاقة بصيغة معدلات النمو. فإذا كانت(u) لا تعاني من مشكلة الارتباط الذاتي. فان(v)ستكون مرتبطة ذاتياً.

قبل نهاية هذه الفقرة لابد من التأكيد على بعض الملاحظات الإجمالية حول الارتباط الذاتي.

فالارتباط الذاتي قد يكون موجباً وقد يكون سالباً أيضاً. وعملياً فان السلاسل الزمنية الاقتصادية تكون غالباً موجبة لان اغلبها إما تتحرك نحو الأعلى أو للأسفل عبر فترة ممتدة. ولا تظهر تغيرات ثابتة نحو الأعلى و الأسفل. والشكل (٦-٥) يوضح ذلك :



(6-4-4) <u>صيغة ماركوف (Markov): الارتباط الذاتي من الرتبة الأولى</u>

إن أكثر أنماط الارتباط الذاتي العملية هي صيغة ماركوف(Markov): الارتباط الذاتي من الرتبة الأولى، والذي يمكن توليده كالآتي:

$$u_t = \rho \; u_{t-1} + \varepsilon_t \qquad . . . \qquad (18-6)$$

حيث أن (ρ) تمثل معامل الارتباط الذاتي : $-1<\rho<1$ والذي يمكن حسابه على وفق التعريف كالاتي:

$$\rho = \frac{\text{cov}(u_t u_{t-1})}{\sqrt{\text{var}(u_t)}\sqrt{\text{var}(u_{t-1})}} = \frac{\text{E}(u_t u_{t-1})}{\text{var}(u_{t-1})}$$

وذلك لان متوسط u صفر، وتباينه متجانس: $\text{var}(u_t) = \text{var}(u_{t-1})$

كما يمكن حسابه بأنه معامل الانحدار في انحدار (u_t)على (u_{t-1})

و ε_t تشير إلى متغير عشوائي يحقق الشروط القياسية للمربعات الصغرى وهي :

متوسط صفري وتباين متجانس، وتباين مشترك صفري :

$\{ E\varepsilon_t = 0 \ , \ \text{var}\,\varepsilon_t = \sigma_\varepsilon^2 \ , \ \text{cov}(\varepsilon_t \varepsilon_{t+s}) = 0$ ***لجميع*** $s \neq 0 \}$ [*] .

و العلاقة (6–18) اختصاراً يرمز لها AR(1) . فهي تمثل علاقة انحدار (u_t) على نفسه بتباطؤ سنة.
وبذلك فان أقصى إبطاء هو (١)

و حيث إن العلاقة (6–18) تتحقق لجميع الفترات الزمنية لذا يمكن إيجاد u_{t-1} , u_{t-2} . . .

$$u_{t-1} = \rho \ u_{t-2} + \varepsilon_{t-1}$$

$$u_{t-2} = \rho \ u_{t-3} + \varepsilon_{t-2}$$

وهكذا بالتعويض المتسلسل عند قيم u_{t-1} , u_{t-2} ,......... . . . يتم الحصول على :

$$u_t = \varepsilon_t + \rho \ \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \cdots + \rho^n \varepsilon_{t-n} + \cdots$$

و هي بذلك توليفة خطية بدلالة ε_t وابطاءاتها المختلفة :

$$u_t = \sum_{s=0}^{\infty} \rho^s \varepsilon_{t-s} \qquad \ldots \qquad (19\text{-}6)$$

وبذلك يمكن الحصول على الآتي باستخدام العلاقة (٦–١٩)

$$E(u_t) = E(\varepsilon_t + \rho \ \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \cdots) = E(\varepsilon_t) + \rho \ E(\varepsilon_{t-1}) + \cdots = 0$$

حيث إن متوسط ε يساوي صفراً .

$$\text{var}(u_t) = E(u_t)^2 = \sigma_\varepsilon^2 + \rho^2 \sigma_\varepsilon^2 + (\rho^2)^2 \sigma_\varepsilon^2 + \cdots$$

$$= \sigma_\varepsilon^2 (1 + \rho^2 + \rho^4 + \cdots)$$

$$= \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

مجموع المتوالية الهندسية :

$$1 + \rho^2 + (\rho^2)^2 + (\rho^2)^3 \cdots = \frac{1}{1 - \rho^2}$$

و كذلك فان التباين المشترك :

$$E(u_t u_{t-1}) = \rho \sigma_u^2 = \frac{\rho}{1 - \rho^2} \sigma_\varepsilon^2$$

وخلاصة القول ان نمط AR(1) للارتباط الذاتي يمكن تلخيصه على وفق الآتي:

(*) ويسمى ايضاً في بعض الدراسات القياسية بالضوضاء البيضاء (white noise)

$$u_t = \rho\ u_{t-1} + \varepsilon_t = \sum_{s=0}^{\infty} \rho^s \varepsilon_{t-s}$$
$$\mathrm{E}(u_t) = 0$$
$$\mathrm{E}(u_t\ u_{t-1}) = \frac{\rho}{1-\rho^2}\sigma_\varepsilon^2 = \rho\sigma_u^2 \quad \ldots \quad (20-6)$$
$$\mathrm{E}(u_t)^2 = \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

وهناك أنماط أخرى للارتباط الذاتي من رتب مختلفة و منها :

AR(2) : 2 nd – order autoregressive scheme وصيغتها:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

AR(3) : 3rd – order autoregressive scheme وصيغتها:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t$$

وهكذا إلى الرتبة PAR(P) ::

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \cdots + \rho_P u_{t-P} + \varepsilon_t$$

ويمكن توليد u_t على وفق ميكانيكية الأوساط المتحركة من الرتب المختلفة:

صيغة الأوساط المتحركة Moving – Average

$MA(1)$: وصيغتها: $u_t = \varepsilon_t + \theta\ \varepsilon_{t-1} \quad ; \quad |\theta| < 1$

$MA(2)$: وصيغتها: $u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$

وهكذا الأوساط المتحركة من الرتبة q :

$MA(q)$: $u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$

فضلاً عن وجود أنماط مركبة لـ u وتسمى الأوساط المتحركة ذاتية الارتباط.

ARMA(P,q) و برتب مختلفة p مرتبطة ذاتياً و q أوساط متحركة. وعلى سبيل المثال فان

$ARMA(1,1)$: تتبع الصيغة التالية: $u_t = \rho_1 u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$

(5-4-6)صفات المقدرات بوجود ارتباط ذاتي بصيغة ماركوف Properties of the estimates in present of autocorrelation

سوف نتطرق في هذه الفقرة إلى دراسة صفات المقدرات بطريقة المربعات الصغرى في حال عدم تحقق الفرض (٥) من فرضيات الانحدار القياسية. ولتوضيح الفكرة يتم الاعتماد على نموذج الانحدار البسيط:

حيث ان:

$$Y = \beta_0 + \beta_1 X + u$$

$$\mathrm{E}\ u_t u_{t-s} \neq 0 \quad .. \quad , s \neq 0$$

وسيتم اعتماد النمط العملي بصيغة ماركوف AR(1) العلاقة (6-19) وعلى وفق افتراضات النموذج فان معلمة الانحدار المقدرة يمكن التعبير عنها :

$$\hat{\beta}_1 = \beta_1 + \frac{\Sigma(X_t - \overline{X})u_t}{\Sigma(X_t - \overline{X})^2}$$

نفترض : $\Sigma(X_t - \overline{X})^2 = A \quad , \quad W_t = X_t - \overline{X}$

إذن:

$$\hat{\beta}_1 = \beta_1 + \frac{W_1}{A}u_1 + \frac{W_2}{A}u_2 + \cdots + \frac{W_n}{A}u_n \quad ... \qquad (21\text{-}6)$$

وبافتراض أن قيم المتغير Xمعطاة ومن ثم $w's$ و A أيضاً معطاة فان $\hat{\beta}_1$ تكون توليفة خطية بدلالة الأخطاء العشوائية. وحيث إن (E(u)=0)، لذا فان المقدرات تبقى غير متحيزة برغم اختلال الفرض (5) ، $\{\mathrm{cov}(u_t u_{t-1}) \neq 0\}$.

و كذلك فان معلمة المقطع الثابت β0:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_0 = (\beta_0 + \beta_1 \overline{X} + \overline{u}) - \hat{\beta}_1 \overline{X}$$

$$\mathrm{E}(\hat{\beta}_0) = \beta_0 + \beta_1 \overline{X} + \mathrm{E}(\overline{u}) - \overline{X}\mathrm{E}(\hat{\beta}_1)$$

$$= \beta_0$$

وبذلك فان وجود نمط الارتباط الذاتي في المتغير العشوائي لعلاقة الانحدار لا يؤدي إلى إيجاد مقدرات متحيزة للمعلمات إي إن المعلمات المقدرة بموجب المربعات الصغرى الاعتيادية تبقى غير متحيزة بالرغم من وجود مشكلة الارتباط الذاتي. غير إن صيغة تباين المعلمات المقدرة لم تعد صحيحة. حيث إن العلاقة (٦-٢١) تنبئ عن وجود ارتباط بين حدود العلاقة وعلى وفق الآتي :

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma x_t^2} + \frac{2\sigma^2}{\Sigma x_t^2}\left[\rho \frac{\sum_1^{n-1} x_t x_{t+1}}{\Sigma x_t^2} + \rho^2 \frac{\sum_1^{n-2} x_t x_{t+2}}{\Sigma x_t^2} + \cdots + \rho^{n-1}\frac{x_1 x_n}{\Sigma x_t^2}\right] \quad \cdots \quad (22-6)$$

وبذلك فان استخدام الصيغة التقليدية لتباين معلمة الانحدار : $(var(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma x^2})$

تكون غير صحيحة في حال وجود ارتباط ذاتي للأخطاء.

وهذا يدلل على إن $var(\hat{\beta}_1)_{OLS} \langle \ (var(\hat{\beta}_1)_{AR(1)}$

وبصيغة أخرى فأن معلمة الانحدار باستخدام المربعات الصغرى عند افتراض AR(1) تكون متحيزة نحو الأسفل أي لا تمتلك أقل تباين. هذا من جهة و من جهة أخرى فان($\hat{\sigma}^2$)هي الأخرى تختلف وهي اكبر من قيمتها الحقيقية و كنتيجة لذلك فان نتائج الاختبارات بالنسبة للمعلمات المقدرة لايوثق بها.

وخلاصة القول فان المعلمات المقدرة (مع وجود ارتباط ذاتي) تبقى خطية وغير متحيزة ولكنها غير كفوءة. وهي النتيجة نفسها التي توصلنا اليها عند وجود عدم التجانس مما يقتضي استخدام طريقة أخرى أكثر كفاءة وهي المربعات الصغرى الموزونة($\hat{\beta}^*$) والتي هي صيغة خاصة من المربعات الصغرى المعممة (GLS).

(6-4-6) الكشف عن مشكلة الارتباط الذاتي. Test of autocorrelation

ان مشكلة الارتباط الذاتي تكون مشكلة مهمة و تحتاج إلى علاج و لكن قبل إي إجراء لابد من استخدام طرائق اختبار ملائمة لإيجاد فيما إذا كانت مشكلة الارتباط الذاتي موجودة أم لا. وفي هذه الفقرة يتم تسليط الضوء على الاختبارات العملية التي في الغالب يتم استخدامها، وحيث إن مشاهدات المتغير العشوائي للمجتمع تكون غير مشاهده، لذا يتم استخدام بدائل عنها وهي مشاهدات الأخطاء (بواقي) علاقة الانحدار.

أولاً: طريقة الرسم Graphical Method

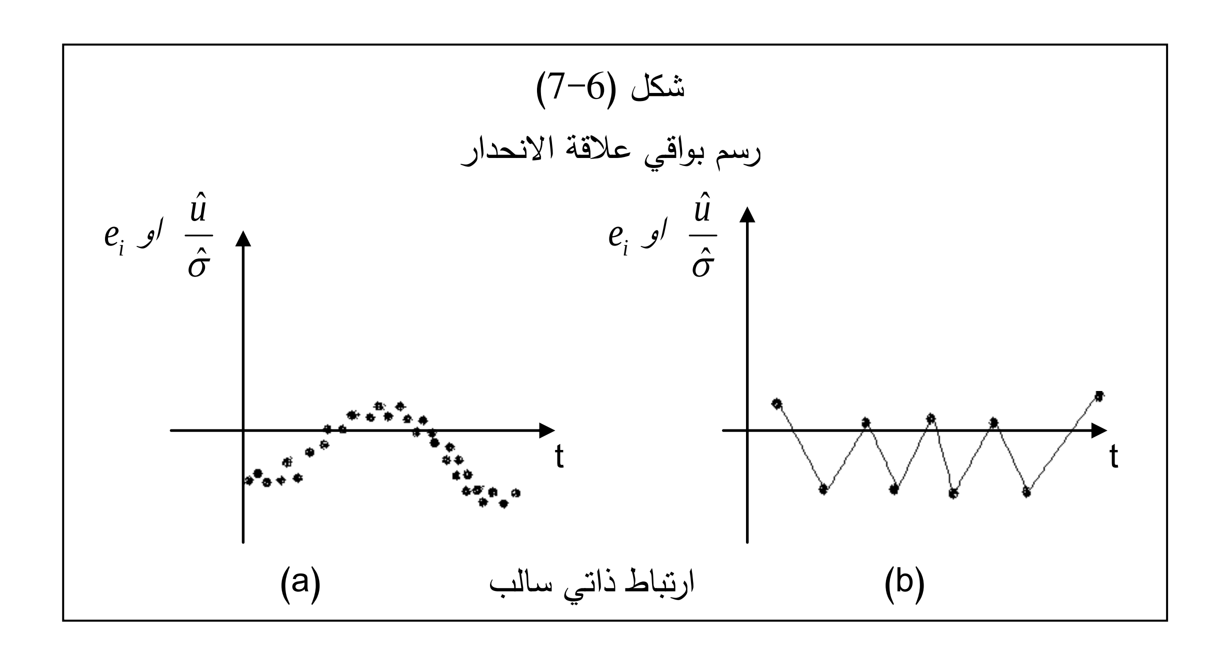
أ) يتم استخدام بواقي علاقة الانحدار (*) ($\hat{u}_t = e_t$) ورسمها ضد الزمن (t) لإعطاء فكرة حول مشكلة الارتباط الذاتي.

ب) و قد تستخدم البواقي المعيارية $\hat{u}/\hat{\sigma}$ (standardized residuals) وترسم ضد الزمن

(*) ان بواقي علاقة الانحدار تعكس خصائص المتغيرات المستقلة وكما تعكس خصائص الأخطاء.

شكل (6-7)

رسم بواقي علاقة الانحدار

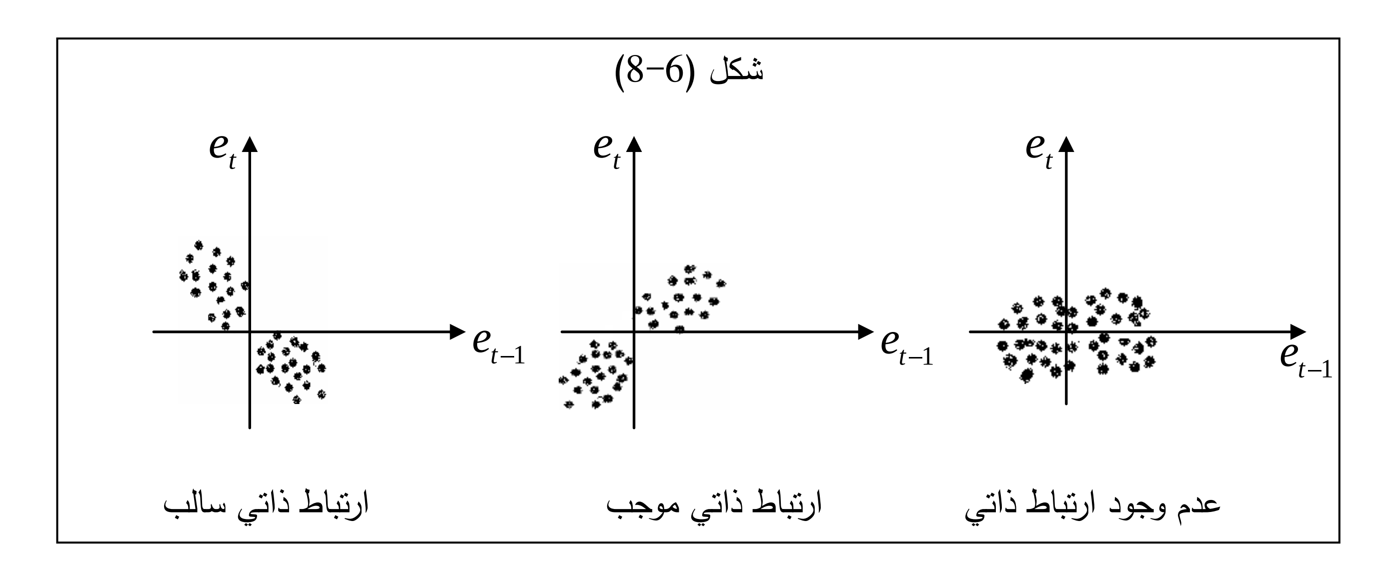


يشير الشكل (a) الى مجموعة من البواقي سالبة تتبعها مجموعة موجبة وهكذا . . . وهذا دليل على عدم وجود العشوائية للبواقي. والشكل يعكس الارتباط الذاتي الموجب، في حين الشكل (b) يظهر تناوب قيم البواقي بالإشارة دليل على وجود ارتباط ذاتي سالب.

ج) كما يمكن رسم البواقي e_t ضد البواقي المتباطئة e_{t-1} فعندما تكون البواقي عشوائية فتكون جميع المشاهدات مبعثرة فإذا اقتصر التبعثر في الربعين الرابع و الثاني فان ذلك يعكس ارتباطاً ذاتياً موجباً للمتغير العشوائي وكذلك إذا كانت جميع مشاهدات الأخطاء مبعثرة في الربعين الأول و الثالث فذلك يدل على وجود ارتباط ذاتي سالب للمتغير العشوائي.

وتجدر الإشارة إلى أن الاختبارات التي تعتمد على الرسم تكون اختبارات تأشيرية وذلك لأنها تعتمد على الحكم الشخصي وعلى خبرة الباحث وخاصة في العينات الصغيرة، لذا لابد من دعمها بالاختبارات الإحصائية والتي تعد صيغها عامة.

شكل (6-8)



ثانياً: طرائق الاختبار الإحصائية :Formal methods

أ)اختبار درين واتسن DW : Durbin –Watson

وهو من أكثر الاختبارات انتشاراً واستخداماً للتحقق من مشكلة الارتباط الذاتي ويعتمد هذا الاختبار على عدد من الافتراضات أهمها :

1- أن يتضمن نموذج الانحدار على مقطع صادي (ثابت).

2- تكون المتغيرات التوضيحية متغيرات غير عشوائية .

3- أن تكون صيغة الارتباط الذاتي للمتغير العشوائي من نمط صيغة ماركوف: الارتباط الذاتي برتبة (1) حصراً ((AR(1)) .

4- أن المتغير العشوائي يتوزع توزيعاً طبيعياً .

5- لا يمتلك نموذج الانحدار أي متغير داخلي متباطئ كأحد المتغيرات التوضيحية .

6- لا توجد مشاهدات محذوفة من النموذج .

و مع تحقق هذه الفروض فان الاحصاءه المستخدمة للاختبار هي :

$$d^* = \frac{\sum_{2}^{n}(e_t - e_{t-1})^2}{\sum_{1}^{n} e_t^2} \quad ... \quad (23-6)$$

حيث أن فرضيات الاختبار :

$$H_0 : \rho = 0 \qquad vs. \qquad H_1 : \rho \neq 0$$

وتقارن القيمة المحسوبة d^* مع القيم الجدولية والتي تعرض في جداول خاصة لعدد مشاهدات (6 – 200) ولعدد متغيرات توضيحية من (1 – 20) وتعرض القيم الجدولية بحدود دنيا(dl) وحدود عليا (du) و لمستويات معنوية 5% أو ١% .

وبموجب صيغة الإحصاءة d^* فان قيمها تمتد من (صفر– 4).

و بتبسيط العلاقة (6–23) يمكن الحصول على النتيجة التالية :

$$d^* = 2\left(1 - \frac{\sum_{t=2}^{n}(e_t\, e_{t-1})}{\sum_{t=1}^{n} e_t^2}\right)$$

ولكن:

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

لذلك فان:

$$. \quad . \quad . \, d^* = 2(1-\hat{\rho}) \qquad (6 -٢٤)$$

وبما ان : $-1 \leq \rho \leq 1$

فان قيمة : $0 \leq d^* \leq 4$

ويكون القرار باستخدام قانون الإبهام (Rule of thumb) كالآتي :

إذا كانت d^* قريبة من (2) فلا يوجد ارتباط ذاتي.

إذا كانت d^* قريبة من الصفر فهناك احتمال وجود ارتباط ذاتي موجب

إما إذا كانت d^* قريبة من ٤ فيدل على احتمال ارتباط ذاتي سالب .

وببساطة يمكن تحديد القرار المناسب لمقارنة القيمة الحسابية d^* ومجالات قبول او رفض فرضية العدم

لايوجد ارتباط ذاتي موجب: H_0 أو

لايوجد ارتباط ذاتي سالب: H_0^*

(dl = 1.077 &du = 1.361

بإتباع المخطط التالي :

المخطط (6-9)

الاحصاءة دربن واتسن d

رفض H_0^* وجود دليل على ارتباط ذاتي سالب	مجال غير حاسم	قبول فرضية: H_0 أو H_0^* أو كلاهما	مجال غير حاسم	رفض H_0 أي وجود دليل للارتباط الذاتي الموجب
4	4−du	4−dl ... 2 ... du	dl	0

ويمكن تلخيص ميكانيكية دربن واتسن بعد تحقق الفروض كالآتي :

1- نجري انحدار Y على X 'S والحصول على البواقي e_i

2- يتم استخدام البواقي لإيجاد الاحصاءة d^* على وفق القانون (٦-٢٣)

3- لعدد المشاهدات المستخدمة nوعدد المتغيرات التوضيحية k تستخدم الجداول الخاصة لإيجاد dl و du عند مستوى دلالة معين .

4- أتباع المخطط (٦-9) لتحديد القرار المناسب.

وجدير بالذكر إن جميع البرامجيات الجاهزة توفر هذه الاحصاءة وعندما تكون قيمة الاحصاءة d^* المحسوبة تقع في المجال غير الحاسم فيمكن استخدام الاختبار المعدل لـ d^* (modified d test) وعند مستوى دلالة $\alpha\%$.

(modified– d) test واختصارا يكتب (m–test)

$$H_0 : \rho = 0 \quad vs. \qquad H_1 : \rho \rangle 0$$

1) لاختبار الفرضية:

اذا $d^* < du$ نرفضH_0 أي يوجد ارتباط ذاتي موجب.

2) كما يمكن اختبار الفرضية: $H_0 : \rho = 0 \quad vs. \quad H_1 : \rho \langle 0$

اذا $4-d^* < du$ نرفضH_0 أي وجود ارتباط ذاتي سالب.

3) أما لاختبار الفرضية: $H_0 : \rho = 0 \quad vs. \quad H_1 : \rho \neq 0$

فيكون القرار على وفق الآتي :

إذا $\begin{cases} d^* \langle du \\ 4 - d^* \langle du \end{cases}$ نرفضH_0 أي وجود دليل إحصائي على ارتباط ذاتي موجب أو سالب.

مثال(6–12) :اختبر فيما اذا كان الارتباط الذاتي لحدود الخطأ موجباً أم لا. إذا علمت أن حجم العينة المستخدمة للانحدار 20 = n. وبتوافر المجاميع التالية :

$$\sum_{2}^{19} (e_t - e_{t-1})^2 = 0.0979 \quad ، \quad \sum_{1}^{20} e_t^2 = 0.1333$$

ومعادلة الانحدار: $\hat{Y} = -1.45 + 0.176X$

الحل: الفرضية المطلوب اختبارها: $H_0 : \rho = 0 \quad vs \quad H_1 : \rho \rangle 0$

$$d^* = \frac{0.0979}{0.1333} = 0.734$$

وحيث أن حجم العينة 20 و $k' = 1$ فان القيم الجدولية :(du = 1.15 , dl = 0.95)

وواضح ان قيمة(d^*) أقل من (dl) فان قاعدة القرار تشير إلى ان الاستنتاج المناسب هو H_1 أي ان حدود الخطأ مرتبطة ذاتياً بصورة موجبة.

مثال (13-6): يمكن استخدام بيانات المثال (3-4) نفسها باستخدام برنامج SPSS لاختبار مشكلة الارتباط الذاتي بصيغة AR(1) .

الحل: بتطبيق SPSS بالاعتماد على بيانات الجدول (3-4) يتم الحصول على الاحصاءة D.W والتي تساعد لاختبار مشكلة الارتباط الذاتي كما في الجدول(6-6)

D.W=1.885

وبمقارنتها مع القيم الجدولية باستخدام مستوى دلالة 5% وحجم العينة (15) مشاهدة:(dl =1.08 ,) du (=1.36) . فتكون $d^* > du$ لذا نقبل فرضية العدم التي تدعم عدم وجود مشكلة الارتباط الذاتي.

ب- اختبار بورش . جود فري (Breusch-Godfrey (BG)test)

وهذا الاختبار يستخدم لاختبار أنماط الارتباط الذاتي العامة:

لنفرض إن :
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \cdots + \rho_P u_{t-P} + \varepsilon_t$$

ε_t : ضوضاء بيضاء.

فرضية العدم:
$$H_0 : \rho_1 = \rho_2 = \rho_3 = \cdots = \rho_P = 0$$

ضد الفرضية البديلة:
$$vs. \quad H_1 : \rho_1 \neq \rho_2 \neq \rho_3 \neq \cdots \neq \rho_P \neq 0$$

فيمكن إجراء خطوات الاختبار كالآتي :

1- انحدار(Y)على جميع ال (X 's) و نحصل على البواقي(e_t).

2- انحدار(e_t) ضد جميع المتغيرات التوضيحية"X 's ( ) و e_{t-1}، e_{t-2}، .....e_{t-p} ونحصل على (R^2) لهذه العلاقة.

3- نحسب الإحصاء BG وفق الآتي :

$$BG = (n-P)R^2 \quad (25\text{-}6) \quad . \; . \; .$$

والتي تتوزع توزيع مربع كاي بدرجات حرية (P)

و يتميز هذا الاختبار بالمميزات التالية :

1- يسمح بوجود متغير داخلي متباطئ Y_{t-1} ، Y_{t-2} ، . . . . كأحد المتغيرات التوضيحية.

2- يسمح باختبار الارتباط الذاتي من نوع AR وبدرجات مختلفة 2،.......P ، كما يمكن اختبار الارتباط الذاتي من نوع الأوساط المتحركة.(ملاحظة:في حالة نمط الارتباط الذاتي (AR(١)) يسمى الاختبار(Durbin' s-m-test) .

3- لكن تبقى رتبة الارتباط الذاتي (P) لا يمكن تحديدها مسبقاً.

مثال (٦-١٤) نفس بيانات المثال (٥ - ١٤).

لاختبار مشكلة الارتباط الذاتي على وفق اختبار Breusch – Godfrey Test(BG)، اعتماداً على البرنامج الجاهز (SPSS):

فيتم في هذه الفقرة انحدار e_t ضد X_1 و X_2 و e_{t-1} وكما موضحة في جدول (٦- ١٠) .

جدول (٦-١٠)

نتائج التحليل الخاصة بالمعلمات

Model	Unstandardized Coefficients		t	Sig.
	B	Std. Error		
(Constant)	0.705	12.77	0.055	0.957
X_1	0.012	0.320	0.067	0.948
X_2	−0.028	0.413	−0.067	0.948
e_{t-1}	−0.0054	0.324	−0.017	0.987

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

فمعادلة التقدير هي:

$$\hat{e}_t = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 e_{t-1}$$

$$= 0/705 + 0.012 X_1 - 0.028 X_2 - 0.00541 e_{t-1}$$

علماً بان احصاءة (بروش – جودفري) تتبع القانون $BG = (n-p)R^2 \sim \chi^2_p$

ولاختبار الارتباط الذاتي AR(1) فان p = 1

جدول (٦-١١)

ملخص النموذج

R	R Square (R^2)	Adjusted R ($\bar{R}^2$) Square	Std. Error of the Estimate
0.025	0.001	−0.299	3.208

المصدر: نتائج برنامج SPSS بالاعتماد على بيانات العينة.

$$BG = (15-1)(0.001) = 0.014$$

$$\chi^2_{c(1,0.05)} = 3.84$$

نرى ان قيمة BG الحسابية أقل من χ^2 الجدولية نقبل فرضية العدم أي ان النموذج لا يعاني من مشكلة الارتباط الذاتي.

ج- اختبار درين – اهج . (testDurbin- h)

في حال احتواء الانحدار على متغير داخلي متباطئ ضمن المتغيرات التوضيحية عند ذلك فان اختبار درين واتسن يصبح لاغياً ويعوض عنه اختبار(h) وتحسب قيمتها على وفق القانون :

$$h = \hat{\rho}\sqrt{\frac{n}{1-n\ \mathrm{var}(\hat{\beta}_2)}} \quad \sim \mathrm{N}(0,1) \quad . . . \quad (26-6)$$

حيث ان β_2 تمثل معلمة المتغير الداخلي المتباطئ بدرجة إبطاء واحدةY_{t-1}.

n : تمثل حجم العينة المستخدمة .

و $\hat{\rho} = 1 - \frac{d^*}{2}$ ، (d^*) تمثل احصاءة درين واتسن

وفي حال كون $(n\ \mathrm{var}(\hat{\beta}_2) \geq 1)$ فان اختبار درين h لايمكن استخدامه ويمكن استخدام الطريقة التالية للاختبار :

1- نجري الانحدار ونستخرج البواقي (e_t)

2- نقدر et ضد e_{t-1} و Y_{t-1} وجميع المتغيرات التوضيحية الأخرى ونختبر معنوية معلمة e_{t-1} فإذا كانت المعلمة معنوية إحصائياً فهذا دليل على وجود مشكلة الارتباط الذاتي.

د- اختبار والصللارتباط الذاتي من الدرجة الرابعة Wallis . AR(4) test.

بافتراض: $u_t = \phi_4 u_{t-4} + \varepsilon_t$ ، حيث أن ε_t : ضوضاء بيضاء

كما يفترض عدم وجود متغيرات عشوائية ضمن المتغيرات التوضيحية، فان اختبار Wallis يختبر الفرضية: $\mathrm{H}_0 : \phi_4 = 0\ vs.\ \mathrm{H}_1 : \phi_4 \neq 0$

وذلك باستخدام الاحصاءه:

$$d_4 = \frac{\sum_{t=5}^{n}(e_t - e_{t-4})^2}{\sum_{t=1}^{n} e_t^2} \quad . . . \quad (27-6)$$

وتقارن القيمة الحسابية d_4 مع جداول Wallis. وتكون على نوعين إما مع وجود مقطع صادي أو بوجود متغيرات وهمية لتعكس حالة الموسمية.

هـ- ويمكن استخدام اختبار (g-statistic): لاختبار فيما إذا كان

$H_1 : \rho \neq 0$ ضد $H_0 : \rho = 0$

وباستخدام الاحصاءة g:

$$g = \frac{\sum_{2}^{n} \hat{v}_t^2}{\sum_{t=1}^{n} e_t^2} \quad \cdots \quad (28-6)$$

حيث ان $\hat{v}_t$: بواقي علاقة الانحدار بصيغة الفرق الأول

مع عدم وجود ثابت في العلاقة .

e_t: بواقي علاقة الانحدار الأصلية (بالمستوى Level).

ويتم استخدام جداول دربن واتسن ايضاً.

مثال (6-15): إذا علمت إن الدخل الشخصي المتاح X_t والاستهلاك الشخصي Y_t والبيانات مقاسة بمليون دينار. اختبر خلو البواقي من الارتباط الذاتي من نوع . AR(1)

Y_t : 199 204 216 218 224 233 238 256 264 270

X_t : 212 214 231 237 244 255 257 273 284 290

الحل:

معادلة التقدير : $\hat{Y} = 7.25 + 0.901X$

تحسب بواقي العلاقة: $e = Y - \hat{Y}$

e_t : 0.74 3.94 0.62 −2.79 −3.09 −4.01 −0.81 2.78 0.87 1.46

Δe_t : 3.2 −3.3 −3.4 −0.3 −0.9 3.2 3.6 −1.9 0.6

تحسب الاحصاءة d*

$$d^* = \frac{60.83}{61.04} = 0.99$$

تقارن مع القيمة الجدولية لعدد مشاهدات(١٠) ومستوى دلالة 1 % وعدد متغيرات مستقلة = 1

dl = 0.604 , du =1.001

من المقارنة نجد ان القيمة الحسابية تقع في مجال غير الحاسم.

- وبالنظر إلى قيم البواقي نجد أن القيم موجبة ثم سالبة ثم موجبة تكون بهيأة (cycle) مما يعطي انطباعاً بان مشكلة الارتباط الذاتي موجودة.

مثال (6- ١٦): إذا علمت إن بواقي علاقة انحدار Y على X هي :

e_t : 0.73 1.05 1.60 2.37 2.14 2.68 0.45 −1.00 −2.23 −2.69 −2.14 −1.37 −0.60 −0.05 −0.51

اختبر خلو النموذج من الارتباط الذاتي بصيغة AR(1) ؟

الحل: $H_0 : \rho = 0 \ vs. \ H_1 : \rho \neq 0$

$$\Sigma(e_t - e_{t-1})^2 = 12.14$$

$$\Sigma e_t^2 = 42.06$$

$$d^* = \frac{12.14}{42.06} = 0.289$$

من جداول دربن واتسن لعدد مشاهدات n = 15 والمتغيرات التوضيحية: k´=1 ولمستوى دلالة 5% يمكن استخراج (dl = 1.077 &du = 1.361)

يتضح من المخطط (6−٩) ان النموذج يعاني من مشكلة الارتباط الذاتي الموجب. على وفق اختبار دربن واتسن.

مثال(6−17): مع توفر لوحة البيانات التالية :

$$\hat{Y}_t = 7.3 - 0.2X_t + 3.5Y_{t-1} \quad , \quad d^* = 1.1229 \quad , \quad t = 1,...,40$$

s.e : (0.04)

اختبر هل إن النموذجيعاني من مشكلة الارتباط الذاتي .

الحل: حيث ان النموذج يحوي متغيراً داخلياً متباطئاً لذا نستخدم اختبار دربن h

$$\hat{\rho} = 1 - \frac{d^*}{2} = 1 - 0.56145 = 0.43855$$

$$h = 0.43855\sqrt{\frac{40}{1-0.064}} = 2.86689 \ \rangle \ 1.96$$

نرفض H_0 أي ان النموذج يعاني من مشكلة الارتباط الذاتي .

مثال (٦ –١٨): لقيم Y و X:

X	1	2	3	4	5
Y	−5	−3	2	7	9

تم الحصول على معادلة التقدير :

$$\hat{Y} = -9.4 + 3.8X$$

اختبر هل إن العلاقة تعاني من الارتباط الذاتي AR(1)

الحل:

تحسب المعلومات الخاصة بالاحصاءة دربن واتسن وكما في الجدول التالي:

جدول حسابات مثال(6−18)

$\hat{Y}$	e_t	e_{t-1}	Δe_t	Δe_t^2	e_t^2
−5.6	0.6				0.36
−1.8	−1.2	0.6	−1.8	3.24	1.44
2	0	−1.2	1.2	1.44	0
5.8	1.2	0	1.2	1.44	1.44
9.6	−0.6	1.2	−1.8	3.24	0.36
Σ				9.36	3.6

ثم نحسب الاحصاءة :

$$d^* = \frac{9.36}{3.6} = 2.6$$

عند مستوى دلالة 5% فأن قيم dudl& : (dl = 0.6 , du = 1.4)

وبذلك فأن القرار :لا توجد مشكلة ارتباط ذاتي بصيغة AR(1) .

ملاحظة:

{عند استخدام الاحصاءة دربن واتسن لابد من التأكد مسبقاً من إن التوزيع للبواقي يكون طبيعياً }

(7-4-6)معالجة الارتباط الذاتي.Remedies of autocorrelation

بعد اكتشاف وجود حدود خطأ مرتبطة ذاتياً . يتم التأكد فيما إذا كانت المشكلة ارتباط ذاتي حقيقي (pure autocorrelation) أم بسبب سوء توصيف النموذج : فقد يكون السبب حذف متغير مهم أو استخدام صيغة دالية غير ملائمة. وللتأكد من ذلك وفي حال السلاسل الزمنية يتم إدخال عنصر الزمن كأحد المتغيرات التوضيحية في المعادلة لمعالجة الاتجاه الزمني في المتغيرات واختبار الارتباط الذاتي، فإذا لم تتخلص المعادلة من الارتباط الذاتي يتم البحث عن متغيرات توضيحية محذوفة و تضمينها في المعادلة وتختبر المعادلة من الارتباط الذاتي. فإذا استمرت النتائج تعلن عن وجود مشكلة الارتباط الذاتي يتم التحقق من الصيغة الدالية المستخدمة فيتم استبدال الخطية بالصيغة التربيعية أو اللوغاريتمية، وإذا استمرت النتائج تؤكد وجود الارتباط الذاتي فهذا يدل على وجود ارتباط ذاتي حقيقي ويتطلب معالجته بتحويل المشاهدات بالشكل الذي يضمن تخليص النموذج من المشكلة.

استخدام المتغيرات المحولة. (Transformation)

بعد التأكد من أن الارتباط الذاتي حقيقي لابد من العمل على تحويل مشاهدات المتغيرات في المعادلة بالشكل الذي يضمن تخليص المعادلة من الارتباط الذاتي للأخطاء.

وعلى سبيل المثال للتوضيح، نفترض معادلة الانحدار الخطي البسيط

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad \ldots \qquad (29\text{-}6)$$

حيث إن $u_t = \rho\, u_{t-1} + \varepsilon_t$ ، علما أن ε_t هي ضوضاء بيضاء.

و ρ معامل الارتباط الذاتي $|\rho| < 1$

فان تحويل المشاهدات يعتمد على قيمة ρ فإذا كانت قيمة ρ معلومة فان تحويل المشاهدات لمتغيرات المعادلة (٦-29) هي:

$$Y_t^* = Y_t - \rho\, Y_{t-1} \qquad \forall \quad t = 2,3, \cdots, n$$

$$X_t^* = X_t - \rho\, X_{t-1} \qquad \forall \quad t = 2,3, \cdots, n$$

$$u_t^* = u_t - \rho\, u_{t-1} = \varepsilon_t \qquad \forall \quad t = 2,3, \cdots, n$$

إما تحويل المشاهدات الأولى فيتم استعمال تحويل (praise winsten) على وفق الآتي:

$$Y_1^* = Y_1\sqrt{1-\rho^2} \qquad , \qquad X_1^* = X_1\sqrt{1-\rho^2}$$

وبذلك فان النموذج المحول :

$$Y_t^* = \beta_0^* + \beta_1^* X_t^* + u_t^* \qquad \ldots \qquad (30\text{-}6)$$

علما بان $\beta_1^* = \beta_1$ و $\beta_0^* = \beta_0(1-\rho)$ و $u_t^* = \varepsilon_t$ وهي ضوضاء بيضاء خالية من الارتباط الذاتي وبعدها نقدر المعادلة المحولة (٦-30) بطريقة المربعات الصغرى الاعتيادية.

طرائق تقدير (ρ)

في حال كون (ρ) غير معلومة فيتم تقديرها بإحدى الطرائق التالية:

1- طريقة الفرق الأول .إي يتم افتراض ρ =١ أو ρ . 1- =وتكون طريقة التحويل مناسبة إذا كان معامل الارتباط الذاتي عالياً جداً. وعندما تكون قيمة الاحصاءة (*d) قليلة جداً، قريبة من الصفر.

ويمكن استخدام اختبار (g–statistic) لاختبار فيما إذا كان

$$H_0 : \rho = 0 \quad \text{ضد} \quad H_1 : \rho \neq 0$$

فعند قبول فرضية العدم فان ذلك دليل على إمكان استخدام (ρ =١) في التحويل.

2- يمكن استخدام تقريب دربن واتسن إعتماداً على العلاقة (٦-24) :

$$\hat{\rho} = 1 - \frac{d^*}{2}$$

وهذا التقريب يكون صحيحاً في حال كون العينة كبيرة.

3- يمكن إجراء انحدار e_t بدلالة e_{t-1} وبدون مقطع ثابت :

$$e_t = \rho\ e_{t-1} + v_t \quad . \ . \ . \quad (31\text{-}٦)$$

وعلى وفق طريقة المربعات الصغرى فان :

$$\hat{\rho} = \frac{\sum_{2}^{n} e_t e_{t-1}}{\sum_{2}^{n} e_{t-1}^2}$$

4- طريقة التكرار (Cochrane–Orcutt iterative)

ويتم استخدام هذه الطريقة سواء كان الارتباط الذاتي من نوع AR(1) أو درجات أعلى .

خطوات الطريقة بافتراض AR(1):

(أ) الحصول على بواقي الانحدار الأصلي (e_t)(استخدام المتغيرات بالمستوى level) .

(ب) إجراء انحدار et بدلالة e_{t-1} وبدون ثابت وكما في (٦-31) ، نحصل على $\hat{\rho}$.

(ج) تستخدم $\hat{\rho}$ لتحويل المشاهدات ويقدر النموذج المحول بطريقة OLS للحصول على المعلمات $\hat{\beta}_0^*$ و $\hat{\beta}_1^*$

$$\hat{Y}^* = \hat{\beta}_0^* + \hat{\beta}_1^* X^*$$

(د) تحسب بواقي علاقة الانحدار بعد التعويض عند قيمة $\hat{\beta}_0^*$ و $\hat{\beta}_1^*$:

$$e^* = (Y - \hat{Y}) = (Y - \hat{\beta}_0^* - \hat{\beta}_1^* X_t)$$

(هـ) تقدر العلاقة: $e_t^* = \hat{\rho}\ e_{t-1}^* + w_t$

$\hat{\rho}$ هي تقدير ρ للجولة الثانية و هكذا تستمر الجولات بشكل متكرر. ويتم التوقف عندما تكون قيمة ρ المتتابعة شبة مستقرة. أي إن التغيرات فيها تكون اقل من (0.01) وعملياً فان الجولة الرابعة تكون هي الجولة الأخيرة.

وجدير بالقول إن اختيار الطريقة لتقدير (ρ) لا يهم كثيراً حيث إن جميعها تعطي مقدرات متسقة.

(٦-4-٨) التنبؤ في حالة وجود حدود خطأ ذاتية الارتباط . Prediction in autocorelated errors

يعد التنبؤ إحدى الاستخدامات المهمة لنماذج انحدار الخطأ ذاتي الانحدار . ففي هذه النماذج يمكن الاستفادة من المعلومات عن حد الخطأ في الفترة الأخيرة (n) للقيام بالتنبؤ بالفترة المستقبلية (n+1) وهذا سيولد تنبؤاً أكثر دقة. وذلك لترابط حدود الخطأ في الفترات المتتالية.

باستخدام نموذج الانحدار الخطي البسيط $Y_t = \beta_0 + \beta_1 X_t + u_t$

و المتغير العشوائي بصيغة: $u_t = \rho\ u_{t-1} + \varepsilon_t$

فنحصل على : $Y_t = \beta_0 + \beta_1 X_t + \rho\ u_{t-1} + \varepsilon_t$

وللفترة (n + 1) $Y_{n+1} = \beta_0 + \beta_1 X_{n+1} + \rho\ u_n + \varepsilon_{n+1}$

وبذلك فان Y_{n+1} تتكون من ثلاثة مركبات :

1- متوسط الاستجابة (القيمة المتوقعة في المدة (n + 1)) : $\hat{Y}_{n+1} = (\hat{\beta}_0 + \hat{\beta}_1 X_{n+1})$

2- حد الخطأ السابق u_n مضاعف بمعامل الارتباط ρ.

3- متغير عشوائي مستقل وبتوقع صفري $E(\varepsilon_{n+1}) = 0$

وبذلك فان التنبؤ للفترة التالية (n + 1) والذي يرمز له بـ $(Y_{f\ n+1})$ وهي تنبؤات شرطية على المشاهدات السابقة Y_n ، Y_{n-1} ، . . . وكذلك شرطية على X_{n+1} والتي في الغالب يتم الحصول عليها وبطريقة الإسقاط

$$Y_{f\ (n+1)} = F_{n+1} = \hat{Y}_{n+1} + \hat{\rho}\ e_{\Omega}$$

وبذلك فان مجال الثقة باحتمال $(1-\alpha)\%$ للقيمة التنبؤية الجديدة تحسب على وفق:

$$Y_{f\ n+1} \pm t_{c(n-3,\frac{\alpha}{2})} \cdot \hat{\sigma}_{Y_{f\ n+1}}$$

ويلاحظ إن درجات الحرية لتقدير $(\rho ، \beta_1 ، \beta_0)$ هي (n-3) لان لدينا (n-1) من البيانات المحولة ما عدا في حال استخدام ρ = 1 .

وبشكل عام فان التنبؤ لأكثر من فترة مستقبلية (h) مثلاً يمكن إيجاده:

$$\hat{Y}_{t+h} = \hat{\beta}_0 + \hat{\beta}_1 X_{t+h} + \hat{\rho}^h e_t$$

مثال(6−18): افترض إن معادلة التقدير : $\hat{Y} = 6.1641 + 1.066X$

و إن قيمة ρ المقدرة : $\hat{\rho} = 0.342$

القيم التنبؤية للمدة (t+1) $\hat{Y}_{t+1} = (6.1641 + 1.066X_{t+1}) + \hat{\rho}\ e_t$

حيث إن: $e_t = Y_t - (6.1641 + 1.066X_t)$

وبذلك فان القيمة التنبؤية للفترة t+1 هي :

$$\hat{Y}_{t+1} = 6.1641 + 1.066X_{t+1} + 0.342(Y_t - 6.1641 - 1.066X_t)$$

القيمة التنبؤية للمدة (t+2) : $\hat{Y}_{t+2} = 6.1641 + 1.066X_{t+2} + (0.342)^2 e_t$

و هكذا . . .

(6−5)المربعات الصغرى المعممة: Generalized least − squares

تمهيد:

عند اختلال الفرضيات (2) أو (3) أو كلاهما فان المتغير العشوائي u لا يكون ضوضاء بيضاء فالمتغير العشوائي لم يعد كروياً (spherical) وان مصفوفة التباين والتباين المشترك للمتغير العشوائي تحدد على وفق الآتي: $\text{var}-\text{cov}(u) = \text{E}\ uu' = \sigma^2\Omega$

حيث ان Ω مصفوفة مربعة موجبة قطعياً (positive definite matrix) بترتيب (n×n) وعليه فان تقدير النموذج الخطي في هذه الحالات لا يتحقق باستخدام طريقة المربعات الصغرى اذ ان (OLS) تعطي مقدرات غير كفوءة ويجب بذلك استخدام طرائق التحويل المناسبة كما اسلفنا في الفقرتين (6−3) و (6−4). وستخصص هذه الفقرة لدراسة طريقة المربعات الصغرى المعممة والتي يرمز لها اختصاراً (GLS) وكما سيتم التأكيد على بعض الحالات الخاصة لهيكل المصفوفة Ω ، هذا فضلاً عن صفات المقدرات التي تعتمد على طريقة (GLS) في حالة كون هيكل مصفوفة Ω معلومة أو غير معلومة.

(6−5−1)المربعات الصغرى المعممة Generalized least – squares.(GLS)

في الانحدار الخطي العام: $Y = X\beta + u$

وبافتراض ان $\text{E}(uu') \neq \sigma^2 I_n$

وانما: $\text{E}(uu') = \sigma^2\Omega$ . . . (6−32)

في حال (Ω) مصفوفة مربعة وموجبة قطعياً.

وبالاعتماد على نتائج التحويل بصيغ المصفوفات يمكن تحويل المتغير العشوائي غير الكروي إلى متغير كروي وذلك بالاعتماد على القاعدة التالية: لاي مصفوفة موجبة قطعياً (Ω) يمكن إيجاد مصفوفة غير شاذة (P) بحيث :

$$PP' = \Omega \qquad \cdots \qquad (33-6)$$

وعليه يتم تحويل النموذج (6–32) وذلك بضرب طرفي العلاقة بـ (P^{-1}) فيتم الحصول على:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u \qquad (34-6) \qquad \cdots$$

وبافتراض:

$$P^{-1}u = u^{*} \ ، \ P^{-1}X = X^{*} \ ، \ P^{-1}Y = Y^{*}$$

فان النموذج المحول:

$$Y^{*} = X^{*}\beta + u^{*} \qquad \cdots \qquad (35-6)$$

حيث ان u^{*} متغير عشوائي بمتوسط صفري وتباينه يحسب على وفق:

$$\operatorname{var}(u^{*}) = \operatorname{var}(P^{-1}u)$$

$$= P^{-1}\operatorname{var}(u)(P^{-1})'$$

$$= \sigma^{2}P^{-1}\Omega(P^{-1})' = \sigma^{2}P^{-1}(PP')(P^{-1})'$$

$$\operatorname{var}(u) = \sigma^{2}P^{-1}PP'(P')^{-1} = \sigma^{2}I_{n} \qquad \cdots \qquad (36)$$

وبذلك فان النموذج المحول تم تخليصه من عدم كروية المتغير العشوائي وأصبحت مصفوفة التباين والتباين المشترك للمتغير العشوائي المحول u^{*} عبارة عن ثابت (σ^{2})مضروبة بمصفوفة الوحدة (I_n)،
وبذلك فان تطبيق المربعات الصغرى على النموذج المحول يعطي مقدرات (BLUE) للمعلمة β.
وبذلك فان معيار المربعات الصغرى المعممة هو تصغير مجموع مربعات بواقي النموذج المحول.
أي:

$$Min \ \Sigma \hat{u}^{*}\hat{u}^{*\prime} \qquad \cdots \qquad (37-6)$$

أو $\qquad e_i^{*} = u_i^{*} = P^{-1}u$

وباستخدام المصفوفات:

$$Min \ \ (P^{-1}e)(P^{-1}e)'$$

أي:

$$Min \ \ P^{-1}ee'(P')^{-1}$$

وبذلك فان المعادلات الطبيعية:

$$X^{*'}X^{*}\beta_{GLS} = X^{*'}Y^{*} \qquad \cdots \qquad (38\text{-}6)$$

$$(P^{-1}X)'(P^{-1}X) \ \ \hat{\beta}_{GLS} = (P^{-1}X)'(P^{-1}Y)$$

$$X'(P')^{-1}P^{-1}X \ \ \hat{\beta}_{GLS} = X'P^{-1'}P^{-1}Y$$

$$X'(PP')^{-1}X \ \ \hat{\beta}_{GLS} = X'(PP')^{-1}Y$$

$$X'\,\Omega^{-1}X \ \ \hat{\beta}_{GLS} = X'\,\Omega^{-1}Y \qquad \cdots \qquad (39\text{-}6)$$

وعليه فان معلمة الانحدار بموجب طريقة GLS تتبع القانون التالي:

$$\hat{\beta}_{GLS} = (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \qquad \cdots \qquad (40\text{-}6)$$

وهذه الصيغة تعرف بصيغة تقدير أتكن (Aitken estimator)

(2−5−6)صفات المقدرات بطريقة المربعات الصغرى المعممة. Properties of the generalized least squares estimates

ان المقدرات التي تتبع الصيغة (6−40) والتي تسمى أيضا ب (مقدرات أتكن) هي المقدرات بطريقة المربعات الصغرى المعممة (GLS) وتتصف بالصفات التالية:

١ − غير متحيزة:

$$\hat{\beta}_{GLS} = (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta+u)$$

$$= (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta + (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}u$$

$$\hat{\beta}_{GLS} = \beta + (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}u \qquad \cdots \qquad (41\text{-}6)$$

$$\therefore \mathrm{E}(\hat{\beta}_{GLS}) = \mathrm{E}\left(\beta + (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}u\right)$$

$$= \beta + (X'\,\Omega^{-1}X)^{-1}X'\Omega^{-1}\mathrm{E}(u)$$

$$= \beta$$

حيث ان: $E(u) = 0$

٢- تركيب خطي بدلالة Y.

حيث ان قيم المتغيرات التوضيحية في المصفوفة X ثابتة للعينة المختارة لذا فان الصيغة (6-40) تعد تركيب خطي بدلالة قيم Y.

٣- مصفوفة التباين والتباين المشترك للمعالم المقدرة بطريقة GLS يمكن اشتقاقها على وفق الاتي: بالاعتماد على العلاقة (6-41):

$$\hat{\beta}_{GLS} = \beta + (X'\,\Omega^{-1}X)^{-1}\,X'\Omega^{-1}u$$

اذن:

$$\hat{\beta}_{GLS} - \beta = (X'\,\Omega^{-1}X)^{-1}\,X'\Omega^{-1}u \quad \cdots \quad (42-6)$$

ولحساب التباين والتباين المشترك يتبع الاتي:

وذلك لان $\hat{\beta}_{GLS}$ غير متحيزة

$$\text{var}-\text{cov}(\hat{\beta}_{GLS}) = \text{E}(\hat{\beta}_{GLS} - \beta)\;(\hat{\beta}_{GLS} - \beta)'$$

وبالتعويض من العلاقة (6-42) يتم الحصول على:

$$= \text{E}\left\{ (X'\,\Omega^{-1}X)^{-1}\,X'\Omega^{-1}u\;u'\Omega^{-1}X\,(X'\,\Omega^{-1}X)^{-1} \right\}$$

$$= (X'\,\Omega^{-1}X)^{-1}\,X'\Omega^{-1}(\text{E}\;uu')\Omega^{-1}X\,(X'\,\Omega^{-1}X)^{-1} \quad \cdots \quad (43\text{-}6)$$

$$= (X'\,\Omega^{-1}X)^{-1}\,X'\Omega^{-1}\sigma^2\Omega\;\Omega^{-1}X\,(X'\,\Omega^{-1}X)^{-1} \quad ، \quad I_n = \Omega\;\Omega^{-1}$$

$$= \sigma^2(X'\,\Omega^{-1}X)^{-1}\,X'\Omega^{-1}X\,(X'\,\Omega^{-1}X)^{-1}$$

$$I_n = X'\Omega^{-1}X\,(X'\,\Omega^{-1}X)^{-1}$$

$$\text{var}-\text{cov}(\hat{\beta}_{GLS}) = \sigma^2(X'\,\Omega^{-1}X)^{-1} \quad \cdots \quad (44\text{-}6)$$

والمصفوفة $(X'\Omega^{-1}X)^{-1}$ مربعة ومتماثلة وموجبة قطعياً وصماء. وعناصر القطر الرئيس تمثل تباين المعلمات المقدرة، اما التباين المشترك بين المعلمات المقدرة فيتمثل بالعناصر غير القطرية.

علماً بان σ^2 تمثل تباين المتغير العشوائي والذي يتم تقديره على وفق القانون:

$$\text{E}(\sigma^2) = \hat{\sigma}^2 = \frac{e^{*'}e^{*}}{n-k-1}$$

حيث ان:

$$e^{*'}e^{*} = (e'\Omega^{-1}e) = TSS_{GLS} - ESS_{GLS}$$

$$= Y'\,\Omega^{-1}Y - \hat{\beta}_{GLS}\,X'\Omega^{-1}Y \quad \cdots \quad (45-6)$$

اي ان:

$$\hat{\sigma}^2 = \frac{Y'\,\Omega^{-1}Y - \hat{\beta}_{GLS}\,X'\Omega^{-1}Y}{(n-k-1)} \qquad . \quad . \quad . \qquad (46-6)$$

وهي تقدير غير متحيز لتباين الخطأ في النموذج (6–32).

تباين المعلمات المقدرة على وفق GLS يكون كفوءاً مقارنة بالمقدرات وفق OLS.

يتضح من المعادلة (6–44) :

$$\text{var}-\text{cov}(\hat{\beta}_{GLS}) = \sigma^2 (X'\,\Omega^{-1}X)^{-1}$$

عندما $\text{E}(uu') = \sigma^2\Omega$

والعلاقة (6–36) أوضحت أن النموذج المحول يحقق الفرضية التي تستلزم المربعات الصغرى ، وعليه فان المقدرات $\hat{\beta}_{GLS}$ هي أفضل مقدرات خطية غير متحيزة لمعلمات النموذج β .

أما في حال استخدام طريقة المربعات الصغرى الاعتيادية للنموذج (6–32) فان مصفوفة التباين والتباين المشترك هي:

$$\text{E}(\hat{\beta}_{OLS}-\beta)\;(\hat{\beta}_{OLS}-\beta)'$$

$$= \text{E}\;\left\{\;(X'X)^{-1}X'u\;u'X\,(X'X)^{-1}\right\}$$

$$= (X'X)^{-1}X'\text{E}(u\;u')X\,(X'X)^{-1}$$

$$\text{var}-\text{cov}(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1}X'\Omega X\,(X'X)^{-1} \qquad . \quad . \quad . \qquad (47-6)$$

ولقياس كفاءة التقدير بموجب GLS نسبة إلى OLS يتم حساب معامل الكفاءة على وفق الصيغة:

$$\text{معامل الكفاءة} = \frac{\text{التباين بطريقة GLS}}{\text{التباين بطريقة OLS}}$$

فإذا كانت قيمة معامل الكفاءة أقل من الواحد الصحيح فان المعلمات المقدرة بطريقة GLS أكثر كفاءة من المقدرات بطريقة OLS. وبمقارنة العلاقة (6–44) و (6–47) يتضح أن معامل الكفاءة اقل من واحد.

وجدير بالذكر أن قيمة المعلمة المقدرة بطريقة GLS على وفق العلاقة (6–40) تعتمد على معرفة عناصر المصفوفة Ω من أجل تحديد المعكوس(Ω^{-1}). وقد تكون عناصر المصفوفة Ω معلومة وغالباً تكون غير معلومة وبذلك ينبغي تقديرها. وسوف نخصص الفقرة التالية لتحديد أهم الأنماط المستخدمة لعناصر مصفوفة Ω.

(3-5-6) أشكال " Ω " الخاصة. Special forms of (Ω)

كما تم توضيحه في الفقرات السابقة أن اختلال الفرضية الكروية للمتغير العشوائي أي : ($E\ uu' \neq \sigma^2 I_n$) فيكون سبب اختلالها من عدد من الأمور، منها وجود مشكلة عدم التجانس أي ان تباين المتغير العشوائي مختلفاً باختلاف المشاهدات. والأمر الآخر المهم هو وجود مشكلة الارتباط الذاتي بين مشاهدات المتغير العشوائي وبإشكاله وهياكله المختلفة. هذا فضلاً عن أمور أخرى هي خارج نطاق مفردات هذا الكتاب وهي الآنية أو النماذج غير المرتبطة ظاهرياً "seemingly unrelated" .

1- اذا المتغيرات العشوائية تعاني مشكلة عدم التجانس (hetroscedasticity) فقط ولكنها غير مرتبطة ذاتياً، وبذلك فان مصفوفة Ω تكون قطرية:

$$\Omega = diag\ \left[\sigma_i^2\right]$$

$$= \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

أما قيم σ_i^2 فهي غالباً تكون غير معلومة وسنتطرق إلى بعض الأمثلة البسيطة في هذا المجال.

مثال (6-19) : افترض n من الشركات و m من الصناعات.

وتمت دراسة انتاج (m) من الصناعات التحويلية مثلاً ولكل منها (n) من الشركات كدالة بدلالة عدد العمال في هذه الشركات الممثلة للصناعات التحويلية وبذلك فان علاقة الانحدار تتبع الآتي:

$$Y_{ij} = \alpha + \beta X_{ij} + u_{ij} \quad (6\text{-}48) \quad . . .$$

i = 1,2, . . . ,m من الصناعات

j = 1,2, . . . ,n من الشركات.

افترض توافر مشاهدات تجميعية حول كل صناعة والتي تحتوي على (n) من الشركات فان المعادلة التجميعية للصناعات:

$$Y_i = \alpha\ n_i + \beta X_i + u_i \quad ; \quad i = 1,2,\ldots,m \quad (6\text{-}49) \quad . . .$$

حيث ان :

$$Y_i = \sum_{j=1}^{n_i} Y_{ij}$$

$$X_i = \sum_{j=1}^{n_i} X_{ij}$$

$$u_i = \sum_{j=1}^{n_i} u_{ij}$$

وبالرغم من افتراض تجانس التباين للمتغير العشوائي في العلاقة (6–48) فان المتغير العشوائي للعلاقة (6–49) يعاني من مشكلة عدم التجانس فهو يتوزع كالآتي: $u_i \sim IID(\ 0,\ n_i\sigma^2)$

$$\sigma_i^2 = n_i\sigma^2$$

حيث ان:

$$\Omega = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_m \end{pmatrix} \qquad \ldots \qquad (50\text{-}6)$$

ويحتاج بذلك تقدير (n) من التباينات وذلك لا يكون ممكناً عند توافر (n) من المشاهدات فقط لذا لتقدير عناصر Ω يكون هناك خياران:

الخيار الأول: المشاهدات المتكررة.

أما الخيار الثاني: فيكون من خلال افتراض معلومات إضافية حول نمط عدم التجانس.

مثال(6–20) : الخيار الاول: المشاهدات المتكررة.

لدراسة الإنفاق الاستهلاكي على الغذاء تم سحب عينة بشكل عشوائي لأسر ضمن فئات دخليه مختلفة (m من الفئات الدخلية). وسحبت (n_i) من الأسر لكل فئة دخلية، وبذلك فان معادلة الانحدار للإنفاق الاستهلاكي على الغذاء:

$$Y_{ij} = \alpha + \beta X_i + u_{ij} \qquad \ldots \qquad (51\text{-}6)$$

$i = 1,2,\ .\ .\ .\ ,m$

$j = 1,2,\ .\ .\ .\ ,n_i$

$$n = \sum_{i=1}^{m} n_i$$ وحجم العينة للنموذج

حيث ان : Y_{ij} الانفاق الاستهلاكي على الغذاء للمشاهدات المتكررة للأسر ذوات نفس الدخل (X_i) .

أي: دالة انحدار الفئة الدخلية الاولى:

$$Y_{1j} = \alpha + \beta X_1 + u_{1j}$$

j = 1,2, . . . ,n_1

n_1 عدد الأسر المسحوبة بشكل عشوائي ذات الدخل X_1.

ودالة انحدار الفئة الدخلية الثانية:

$$Y_{2j} = \alpha + \beta X_2 + u_{2j}$$

j = 1,2, . . . ,n_2

n_2 عدد الاسر المسحوبة بشكل عشوائي لفئة الدخل X_2.

وهكذا . . .

$$Y_{mj} = \alpha + \beta X_m + u_{mj}$$

j = 1,2, . . . ,n_m

n_m عدد الأسر المسحوبة بشكل عشوائي لفئة الدخل X_m.

وبذلك فان حجم العينة المسحوبة هي $n = (n_m + . . . + n_2 + n_1)$

كما ان u_{ij} متغيرات عشوائية تتوزع توزيعاً مستقلا بمتوسط صفر وتباين σ_i^2 يعتمد على حجم العينة في الفئة الدخلية المعينة. ويمكن تقديره لـ (m) من الفئات الدخلية بطريقتين:

الطريقة الأولى: تحسب على وفق القانون:

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{(n_i - 1)} \quad . . . \quad (52-6)$$

حيث:

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$$

أما الطريقة الثانية فتحسب على وفق الآتي:

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i} {e_{ij}}^2}{n_i} \quad . . . \quad (53-6)$$

حيث e_{ij} تمثل بواقي انحدار المربعات الصغرى. $e_{ij} = Y_{ij} - \hat{\alpha} - \hat{\beta} X_i$

والتقديرين (6−52) و (6−53) تقديرات متسقة (consistent) لـ σ_i^2.

الخيار الثاني: توقع نمط عدم التجانس على وفق المشكلة التي تتم معالجتها.

مثال (6−21): دراسة بيانات مقطعية للربح بدلالة المبيعات

نتوقع ان الشركات الكبيرة تمتلك مصادر تحويل اكبر ويمكنها افتراض كميات اكبر او استثمار اكبر او تكون خسارتها او ربحها اكبر من الشركات الصغيرة. وعليه فان نمط عدم التجانس يكون مرتبطاً مع حجم الشركة والذي يعكس المتغير المستقل كأن يكون المبيعات او أي متغير آخر يقيس الحجم.

وبشكل عام يمكن تقدير σ_i^2 على وفق الآتي:

- صيغة حاصل الضرب

$$\sigma_i^2 = \sigma^2 Z_i^{\delta} \quad . \; . \; . \quad (54\text{−}6)$$

(δ) معلمة غير معلومة.

ويمكن تقدير قيمة (δ) على وفق الاختيارات الاتية : $\delta = 0, 0.1, 0.2, \cdots, 4, \cdots$

وان Z_i يمكن ان تكون أحد المتغيرات التوضيحية (X_i) أو $\hat{Y}_i = E(Y_i)$

وهنا نحتاج أن نقدر σ^2 و δ فقط عوضاً عن تقدير (n) من التباينات المختلفة (σ_i^2)

أو بصيغة الجمع:

$$\sigma_i^2 = a + b_1 z_{1i} + b_2 z_{2i} + \cdots + b_r z_{ri} \quad . \; . \; . \quad (55\text{−}6) \;, \; r < n$$

٢− اذا المتغير العشوائي يتبع نمط الارتباط الذاتي (Autocorelation) فقط وتباين المتغير العشوائي متجانس . فان مصفوفة Ω تعتمد على نمط الارتباط الذاتي ويمكن تحديد عناصر مصفوفة Ω للأنماط التالية:

(أ) اذا كان المتغير العشوائي يتبع نمط AR(1) . وبذلك وكما تم توضيحه في المبحث السابق فان:

$$u_t = \rho \, u_{t-1} + \varepsilon_t \qquad , t = 1, \ldots, n$$

بافتراض $|\rho| \rangle 1$ ، وان ε_t ضوضاء بيضاء، وبذلك فان $\varepsilon_i \sim IID(\,0,\; \sigma_\varepsilon^2)$ ، والتباين المشترك:

$$\text{cov}(u_i u_{t-s}) = \rho^s \sigma_u^2$$

علماً بان: $\sigma_u^2 = \dfrac{\sigma_\varepsilon^2}{1-\rho^2}$

وبذلك فان مصفوفة Ω تحدد عناصرها على وفق الآتي : $E(u\,u') = \sigma^2 \Omega$

$$\Omega = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \quad . \; . \; . \quad (56\text{−}6)$$

(ب) اذا كان المتغير العشوائي يتبع نمط MA(1) .

أي ان:

$$u_t = \varepsilon_t - \theta\, \varepsilon_{t-1}$$

حيث ان ε_t متغير عشوائي ضوضاء بيضاء كما أسلفنا.

وحيث ان:

$$\text{var}-\text{cov}(u) = \begin{pmatrix} \text{var}(u_1) & \text{cov}(u_1u_2) & \text{cov}(u_1u_3) & \cdots & \text{cov}(u_1u_n) \\ & \text{var}(u_2) & \text{cov}(u_2u_3) & \cdots & \text{cov}(u_2u_n) \\ & & & \ddots & \vdots \\ & & & & \text{var}(u_n) \end{pmatrix}$$

أي ان عناصر القطر تمثل التباين.

وعناصر المثلث العلوي تماثل عناصر المثلث السفلي وتمثل التباين المشترك. وتحسب كالآتي:

$$\begin{aligned}\text{var}(u_i) = \text{E}(u_t)^2 &= \text{E}(\varepsilon_t - \theta\, \varepsilon_{t-1})^2 \\ &= \text{E}(\varepsilon_t^2 - 2\theta\, \varepsilon_t \varepsilon_{t-1} + \theta^2 \varepsilon_{t-1}^2) \\ &= \sigma_\varepsilon^2 + \theta^2 \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 (1+\theta^2)\end{aligned}$$

$$\begin{aligned}\text{cov}(u_i u_{t-1}) &= \text{E}\{(\varepsilon_t - \theta\, \varepsilon_{t-1})\ (\varepsilon_{t-1} - \theta\, \varepsilon_{t-2})\} \\ &= \text{E}\{\ \varepsilon_t \varepsilon_{t-1} - \theta \varepsilon_t \varepsilon_{t-2} - \theta\, \varepsilon_{t-1}^2 + \theta^2 \varepsilon_{t-1} \varepsilon_{t-2}\} \\ &= -\theta\, \sigma_\varepsilon^2\end{aligned}$$

$$\text{cov}(u_i u_{t-2}) = \text{E}\{(\varepsilon_t - \theta\, \varepsilon_{t-1})\ (\varepsilon_{t-2} - \theta\, \varepsilon_{t-3})\} = 0$$

وهكذا بالنسبة للتباين المشترك لابطاءات أكثر من (2) تساوي صفراً.

اذن تكون عناصر Ω كالآتي:

$$\Omega = \begin{bmatrix} 1+\theta^2 & -\theta & 0 & 0 & \cdots & 0 \\ -\theta & 1+\theta^2 & -\theta & 0 & \cdots & 0 \\ \rho^2 & -\theta & 1+\theta^2 & -\theta & \cdots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -\theta & 1+\theta^2 \end{bmatrix}$$

أسئلة الفصل السادس:

س1: 1. وضح المقصود إحصائياً بأن متوسط المتغير العشوائي يساوي صفراً.
2. ناقش احتمالات عدم تحقق الفرض في اعلاه موضحاً تأثير ذلك فى نتائج التقدير.

س2: 1. وضح المقصود احصائياً بان متوسط المتغير العشوائي يساوي صفراً.
2. ناقش تأثير اختلال الفرض على نتائج التقدير.
3. اذكر أهم الاختبارات المستخدمة للكشف عن تحقق الفرض.
4. اذكر أهم الطرائق لتجاوز مشكلة عدم تحقق التوزيع الطبيعي.
5. اذكر مثالاً عملياً يتوضح به عدم تحقق التوزيع الطبيعي.

س3: صحح العبارات الخاطئة ان وجدت.

١. عند وجود متغير داخلي متباطئ في معادلة الانحدار كأحد المتغيرات التوضيحية يجعل المقدرات متحيزة وغير كفوءة.
٢. في حالة الترابط بين u_i و X_i سالباً فان المقطع الصادي المقدر يكون متحيزاً الى الاسفل والميل يكون متحيزاً الى الاعلى والعكس صحيح في حالة وجود ترابط موجب.
٣. المختبر JB يتوزع توزيع كاي وبدرجات حرية k وهي عدد المعلمات في النموذج.
٤. ان اختلال التوزيع الطبيعي يتلازم مع فرضية استقلال مشاهدات المتغير العشوائي عن بعضها.
٥. اتباع نماذج اخطاء التعلم تولد تحيزاً للمعلمات المقدرة.
٦. في دراسة المقاطع العرضية دليلاً على المتغير العشوائي له تباينات متجانسة.
٧. يتولد سوء التوصيف في نموذج الانحدار عند وجود تغيرات هيكلية.
٨. تحليل البواقي احدى الطرائق المستخدمة لتقدير معلمات الانحدار اذا كانت البواقي غير متجانسة.
٩. اختبار بارك يستخدم لمعرفة نمط عدم تجانس تباين الخطأ الذي يسهم في تحويل المشاهدات كطريقة لعلاج المشكلة.
١٠. اختبار (BPG) بروش–بيجن–جودفري يتوزع توزيعاً طبيعياً، في حين اختبار white يتوزع على وفق مربع كآي.

س4: وضح أهم الحالات التي تولد سوء التوصيف في انموذج الانحدار، وضح أثر ذلك فى نتائج التقدير.

س5: ناقش أهم الطرائق المستخدمة للتحري عن مشكلة عدم تجانس تباين اخطاء الانحدار.

س6: وضح عناصر المصفوفة Ω في الحالات التالية:

١. مشاهدات (Y_i) عبارة عن متوسطات.
٢. تباين المتغير العشوائي يتناسب مع مربع (توقع Y).
٣. $E(u_i^2) = \sigma^2 X_i^2 \qquad \forall \quad i$
٤.

$$|\hat{e}_i| = 2.1 + 0.5\frac{1}{X}$$
$$s.e \quad : \qquad (0.03)$$

٥. المتغير العشوائي يكون بصيغة ماركوف.
٦. المتغير العشوائي يكون بصيغة الاوساط المتحركة من الرتبة الاولى.

س7: اذكر صفات المقدرات بطريقة OLS في الحلات التالية:

١. وجود متغير توضيحي عشوائي.
٢. تباين البواقي غير متجانس.
٣. وجود ارتباط بصيغة ماركوف.

س٨: وضح بالرسم الحالات التالية:

١. عدم تجانس تباين الخطأ.
٢. المتغير العشوائي مرتبط ذاتياً.

س٩:

أ) أهم الافتراضات لاستخدام الاحصاءة دربن واتسن.
ب) الاختبارات البديلة مع عدم تحقق بعض من الفروض.
ج) أهم الطرائق المتفق على استخدامها لتقدير معامل الارتباط الذاتي.

س١٠: وضح مفهوم المتغير العشوائي الكروي spherical disturbance .

س١١:

أ) صفات المقدرات بطريقة OLS و GLS في الحالات التالية:

1. $\text{var}(u) = \sigma^2 X$

2. $u_t = 0.9u_{t-1} + \varepsilon_t$

ب) حدد تباين المعلمات المقدرة بطريقتي OLS و GLS لكل من الحالات (1) و(2)

ج) اذكر الصيغة المناسبة لتقدير تباين الخطأ على وفق كل من OLS و GLS .

د) احسب كفاءة التقدير.

الـفـصـل الـسـابـع

تمهيد:

يتضمن هذا الفصل مناقشة فرضيات التحليل التي تخص البيانات وهي الفرضيات 2، 6، 7، 8، 10.وبذلك فإن فقرات هذا الفصل تم تخصيصها لمناقشة الاختلال في هذه الفرضيات تباعا:

(1-7) <u>اختلال الفرض (2) :بمعنى أن المتغيرات التوضيحية عشوائية.</u>

الفرض (2) ينص على ان المتغيرات التوضيحية (X 's) متغيرات مستقلة أي ان مشاهداتها ثابتة للعينة المختارة. أي ان كل من هذه المتغيرات تكون غير عشوائية.

وعند نقض هذه الفرضية أي اذا كان أحد المتغيرات التوضيحية أو جميعها متغيراً عشوائياً.

وتتأكد هذه عند الحالات الثلاث التالية:

أ- Xمتغير عشوائي وغير مرتبط بحد الاضطراب العشوائي (u) وهذا ما تنص عليه الفرضية (6). وبذلك تبقى المربعات الصغرى والإمكان الأعظم تولد مقدرات غير متحيزة وخطية وكفوءة أي أنها (BLUE) وبذلك لاتوجد مشكلة قياسية في تقبل النتائج.

ب- عندما X متغير عشوائي ظاهرياً. وهذه الحالة تبرز عندما يكون أحد المتغيرات التوضيحية هو متغير معتمد متباطئ مثلاً Y_{t-1} أو Y_{t-2} وهكذا.

وكمثال على ذلك عند دراسة الإنفاق الاستهلاكي Y_t كدالة بدلالة العادات الاستهلاكية كأن تكون الدراسة هي تقدير الانفاق الاستهلاكي للتدخين فتصاغ الدالة Y_t دالة بدلالة العادات الاستهلاكية والتي يمكن تمثيلها بالمتغير)(Y_{t-1} وهو متغير الاستهلاك في المدة السابقة وبذلك فان دالة الانحدار:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t \quad ; \quad t = 2,3,\cdots$$

ومعلوم ان Y_t مرتبط مع u_t . والسؤال هنا هل ان Y_{t-1}مرتبط مع u_t ؟

ومع تحقق فرضية ان u_t غير مرتبط ذاتياً أي:
$$\operatorname{cov}(u_t u_{t-s}) = 0$$

وبذلك فان u_t غير مرتبط مع Y_{t-1}وهذا ما نعني به غير مرتبط ظاهرياً .

ان u_t مرتبط مع Y_t ولكنه غير مرتبط مع Y_{t-1}.

غير ان تطبيق OLS في التقدير يعطي مقدرات تفقد خاصية عدم التحيز.

حيث ان:

$$\hat{\beta} = \frac{\Sigma y_t y_{t-1}}{\Sigma y_{t-1}^2} = \beta + \frac{\Sigma y_{t-1}\, u_t}{\Sigma y_{t-1}^2}$$

ولكن حقيقة الأمر ان:

$$E\frac{\Sigma y_{t-1}\ u_t}{\Sigma y_{t-1}^2} \neq \frac{E(\Sigma y_{t-1}\ u_t)}{E(\Sigma y_{t-1}^2)}$$

لان $y_{t-1} = (Y_{t-1}\bar{Y})$ وان $\bar{Y}$ تتضمن Y_t حيث $\bar{Y} = \frac{\Sigma Y_t}{n}$

وحيث $(EY_t\ u_t \neq 0)$ وبذلك $E(\hat{\beta}) \neq \beta$

غير ان التحيز مع كبر حجم العينة يزول وعليه فان المعلمة β المقدرة تكون متسقة بمعنى مع وجود متغير داخلي متباطئ فان المقدرات بطريقة OLS تكون متحيزة ولكنها متسقة وكفوءة وخطية مع كبر حجم العينة.

(ج) عندماتكونX متغيراً عشوائياً وبالوقت نفسه مرتبط بحد الاضطراب العشوائي u .

وعندها فان المربعات الصغرى تكون متحيزة وكذلك غير متسقة أي ان مقدار التحيز يستمر حتى مع كبر حجم العينة وبذلك تصبح المربعات الصغرى غير مرغوبة ويستعاض عنها بطرائق اخرى.

وهناك عدة حالات يتم انتهاك هذه الفرضية فيها ومن أبرز هذه المسائل هي:

i. وجود خطأ في القياسات error in measurement .

ii. حالة المتغيرات الداخلية المتباطئة عندما يكون حد الخطأ مرتبط ذاتياً.

iii. حالة المعادلات الآنية simultaneous equations .

ومناقشة هذه الحالات هي خارج منهج الكتاب. غير انه من المفيد معرفة أن استخدام المربعات الصغرى في حالة وجود ترابط موجب بين u_i و X_i فان المقطع الصادي المقدر يكون متحيز سفلي (under - estimated) . ومعلمة الانحدار(الميل) يكون متحيزاً علوياً (over- estimated).

أما في حالة كون الترابط بين u_i و X_i سالباً فان المقطع الصادي المقدر يكون متحيزاً للأعلى (over- estimated) . ويكون الميل متحيزاً تحيزاً سفلياً (under- estimated). علماً بان مقدار التحيز يستمر مع كبر حجم العينة فتبقى المعلمات غير متسقة.

الفرضية (6): اذا $X's$ تكون عشوائية فان حد الاضطراب u و $X's$ تكون مستقلة أو غير مرتبطة.

$$E(X'u) = 0$$

وان اختلالها يكون اما $X's$ عشوائية ظاهريا أو عشوائية وارتباطها مع حد الاضطراب يولد مشكلات قياسية وكما تم توضيحها في الفقرة أعلاه(٧ –1) (ب) و (ج).

الفرضية (7): عدد المشاهدات (n) يجب ان يكون اكبر من عدد المتغيرات التوضيحية (k) أي kn > .

واختلال الفرضية سوف يجعل درجات الحرية سالبة وهذا غير منطقي وبذلك يجعل الاختبارات الاحصائية غير موثوق بها وغير ممكنة التطبيق.

الفرضية (8): وجود تغيرات كافية وملموسة في مشاهدات كل متغير توضيحي.

ان اختلال هذه الفرضية لا يؤثر فى صفات المقدرات بطريقة المربعات الصغرى. غير ان الانحراف المعياري للمقدرات سيكون اكبر نسبة الى معلماتها المقدرة وبذلك فان قيم (نسبة t) ستكون قليلة، وبذلك ستجعل مساهمة المتغير التوضيحي المعني(الذي تكون التغيرات في مشاهداته غير كافية)في مجموع المربعات المشروحة غير واضحة.

الفرضية (10). عدم وجود علاقة تامة بين المتغيرات التوضيحية $X's$ ، وفي حال نقض هذه الفرضية تظهر مشكلة ما يسمى التعدد الخطي.

<u>(2-7)التعدد الخطي (Multicollinearity)</u>

<u>المفهوم</u>: المصطلح الانكليزي "Multicollinearity " مصطلح مركب من ثلاثة مقاطع:

(Multi) : ومعناها متعدد و(co) والذي يقابله بالعربية تداخل أو تناظر، و (linearity) : وتعني الخطية. وقد اختلفت الترجمات العربية لهذا المصطلح منها : الارتباط الخطي المتعدد، الامتداد الخطي المتعدد، تعدد العلاقات الخطية، التشابك أو التداخل الخطي أو التشابك الخطي المتعدد، وسوف نستخدم تداخل التعدد الخطي واختصاراً " **التعدد الخطي** " ويعود هذا المصطلح الى النرويجي (Ragnar Frisch 1934) الذي وجه الاهتمام الى هذه الظاهرة في معرض تحليله لبيانات السلاسل الزمنية. فوجد ان المتغيرات الاقتصادية التوضيحية متناغمة عبر الزمن مما يفقد بعضها المقدرة التفسيرية في النموذج .

والمشكلة التي نحن بصدد توضيحها توجد في نماذج الانحدار الخطي المتعدد حصراً ولا تعاني منها مشكلات الانحدار البسيط ، فهي تنشأ بسبب تداخل (تشابك) بين المتغيرات التوضيحية (التفسيرية) ذاتها. كما ان الفرضية التي ترتبط بوجود قدر كافٍ من التغيرات في قيم المتغيرات التوضيحية وان عدد المشاهدات يكون اكبر من عدد المتغيرات التوضيحية هي فرضيات متممة للتعدد الخطي .

ويمكن تحديدها بالفرضية :

{ رتبة مصفوفة المعلومات(X) تكون تامة من ناحية الاعمدة.}

وبالرموز :

$$\rho(X) = k + 1 < n \quad \ldots \quad (1\text{-}7)$$

حيث ان :n حجم العينة المستخدمة في الانحدار.

k : عدد المتغيرات التوضيحية في المعادلة .

(k+ 1): عدد المعلمات في المعادلة بضمنها المقطع الثابت.

والمفهوم يتضمن " التعدد الخطي التام "(perfect multicolinearity)" كما يتضمن التعدد الخطي شبه التام (semi- perfect multicolinearity)

ان التعدد الخطي التام يتحقق اذا كان واحد أو اكثر من المتغيرات التوضيحية (المستقلة) توليفة خطية تامة من المتغيرات الأخرى كالآتي:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0$$

λ َ S: ثوابت ليست جميعها أصفاراً .

اما التعدد الخطي شبه التام فيتحقق اذا ارتبطت المتغيرات التوضيحية أو المستقلة ببعضها ارتباطاً قوياً ولكن غير تام :

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k + e_i = 0$$

e_i: متغير عشوائي.

وفي حالة انعدام مشكلة التعدد الخطي فان المتغيرات التوضيحية تسمى متغيرات متعامدة (orthogonal) أي ان كل متغير توضيحي يؤثر في المتغير التابع بطريقه منفصلة تماماً. وبذلك تكون نتائج التقدير في الانحدار المتعدد متطابقة تماماً مع نتائج التقدير في الانحدار البسيط حيث ان مصفوفة فيشر(X´X) تكون قطرية لان معامل الارتباط بين المتغيرات التوضيحية معدوم $(r_{X_i X_j} = 0)$ أو ان الارتباط المشترك بين المتغيرات التوضيحية يسأوي صفراً .

$$\sum_{t=1}^{n} (X_{ti} - \bar{X}_i)(X_{tj} - \bar{X}_j) = 0 \qquad \forall \qquad i \neq j = 1,2.\cdots,k$$

(7- 3)<u>أسباب وجود التعدد الخطي Reasons of multicollinearity</u>

في حاله التعدد الخطي التام فان أهم مسببات وجود المشكلة يمكن تحديده على وفق الآتي .

1) ان التعريف الخاطئ للمتغيرات وخاصة (الوهمية)* ينذر بوجود مشكلة التعدد الخطي التام وهي ما يسمى بمصيدة المتغيرات الوهمية " Dummy variable trap "

2) ان تكون قيم احد المتغيرات ذات تغيرات قليلة ومنها ينشأ ترابط بين العمود الذي يمثل قيم ذلك المتغير التوضيحي مع العمود الأول في مصفوفة المعلومات X والذي يمثل قيم المقطع الثابت في علاقة الانحدار .

3) استخدام وحدتي قياس مختلفة للمتغير التوضيحي نفسه.

(*) سيتم توضيح ذلك في الفصل الخاص بالمتغيرات الوهمية (Dummy variables).

أما الأسباب التي تولد مشكلة التعدد الخطي شبه التام فهي كثيرة ومتعددة نلخص أهمها.

1) ميل المتغيرات التوضيحية للتغير سوياً عبر الزمن خاصة في دراسات السلاسل الزمنية.
2) وجود متغيرات توضيحيه متباطئة (متخلفة زمنياً).
3) وجود مشاهدات شاذة (متطرفة).
4) قد يكمن السبب في طريقة جمع البيانات للعينة. فمثلاً عند جمع عينة من مجال محدود لقيم المتغيرات التوضيحية في المجتمع.
5) وجود قيود في المجتمع الذي تؤخذ منه العينة. على سبيل المثال عند دراسة الطلب على الكهرباء بدلالة الدخل وحجم السكن، يوجد قيد في المجتمع لهذه العوائل، فالعوائل ذات فئات الدخل العالي تمتلك بيوتاً كبيرة بنسبة الى فئات الدخل المنخفضة.
6) قد توجد مشكلة التعدد الخطي شبه التام عندما يكون عدد المتغيرات التوضيحية كبيراً (مقارنة بعدد المشاهدات للعينة).

(7- 4) التقدير بواسطة المربعات الصغرى Least square estimates

(7-4-1) التقدير في حالة مشكلة التعدد الخطي التام Perfect multicollinearity

في هذه الحالة فان المعلمات المطلوب تقديرها لايمكن إيجادها بشكل منفصل وان انحرافها المعياري يكون كبيراً. ويمكن توضيح هذه الحالة كالآتي: $y = \beta_2 x_2 + \beta_3 x_3 + u$

افترض: $x_3 = \lambda x_2$ ، $\lambda \neq 0$

$$\hat{\beta}_2 = \frac{(\Sigma y x_2)(\Sigma x_3^2) - (\Sigma y x_3)(\Sigma x_2 x_3)}{(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2 x_3)^2} = \frac{\Sigma y x_2 \Sigma \lambda^2 x_2^2 - \lambda^2 \Sigma y x_2 \Sigma x_2^2}{(\Sigma x_2^2)\lambda^2(\Sigma x_2^2) - \lambda^2(\Sigma x_2^2)^2} = \frac{0}{0}$$

$$\hat{\beta}_3 = \frac{(\Sigma y x_3)(\Sigma x_2^2) - (\Sigma y x_2)(\Sigma x_2 x_3)}{(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2 x_3)^2} = \frac{\lambda \Sigma y x_2 \Sigma x_2^2 - \lambda \Sigma y x_2 \Sigma x_2^2}{(\Sigma x_2^2)\lambda^2(\Sigma x_2^2) - \lambda^2(\Sigma x_2^2)^2} = \frac{0}{0}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \frac{0}{0} \\ \frac{0}{0} \end{pmatrix} \Rightarrow$$

أي لايمكن تحديد قيمة $\hat{\beta}_2$ و $\hat{\beta}_3$ على وفق القانون بشكل منفصل

ولكن يمكن التعويض في النموذج كالآتي:

$$y = \beta_2 x_2 + \beta_3 \lambda x_2 + u$$
$$= (\beta_2 + \beta_3 \lambda) x_2 + u$$
$$= \alpha \ x_2 + u \quad ; \quad \alpha = \beta_2 + \beta_3 \lambda$$
$$\hat{\alpha} = \frac{\Sigma x_2 y}{\Sigma x_2^2} = \hat{\beta}_2 + \lambda \hat{\beta}_3$$

معادلة واحدة بمجهولين $(\hat{\beta}_3$ و $\hat{\beta}_2)$ $\Leftarrow$ لا يمكن حلها لإيجاد $\hat{\beta}_2$ و $\hat{\beta}_3$ بشكل منفرد .

وبذلك نستنتج انه:

في حالة التعدد الخطي التام لايمكن إيجاد حل وحيد لكل معلمة من معلمات النموذج . وإنما يمكن إيجاد تقدير التراكيب الخطية من هذه المعلمات.

(2-4-7) التقدير في حالة وجود تعدد خطي عالي (شبه تام) Semi multicollinearity

في الواقع العملي لا يوجد ارتباط خطي تام بين المتغيرات التوضيحية وخاصة في دراسات السلاسل الزمنية وانما تكون العلاقة بين المتغيرات شبه تامة :

افترض : $x_{3t} = \lambda x_{2t} + v_t$

λ : ثابت

v : متغير عشوائي غير مرتبط مع X_2. بمعنى ان علاقة الارتباط الخطي بين المتغيرين X_2 و X_3 $(0 \langle \ r_{X_2X_3} \langle \ 1)$ لا تسأوي واحدا.

وبذلك ان تقدير المعلمات يصبح ممكناً ولكن هناك آثاراً أخرى يمكن توضيحها.

ان استخدام المربعات الصغرى يولد مقدرات غير متحيزة ومتسقة والانحراف المعياري للخطأ يتم تقديره بشكل صحيح ولكن قيمته تكون كبيرة.

$$\hat{\sigma}^2 = \frac{\Sigma e_i^2}{n - k - 1}$$

– اذ ان عدم التحيز بمفهومه النظري (ان قيمة المعلمة المقدرة في المتوسط تكون مسأوية الى قيمة المعلمة في المجتمع ، بشرط تكرار العينة) يكون متحققاً. غير ان ذلك لا يعني شيئاً حول صفات المقدرات في أي عينة من عينات المجتمع.

. كما ان الانحراف المعياري للمعلمات المقدرة يكون صحيحاً .

افترض علاقة الانحدار: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$

في الانحدار يحسب التباين والتباين المشترك على وفق:

$$
\left.
\begin{aligned}
\operatorname{var}(\hat{\beta}_1) &= \frac{\hat{\sigma}^2}{\Sigma x_1^2 (1 - r_{12}^2)} \\
\operatorname{var}(\hat{\beta}_2) &= \frac{\hat{\sigma}^2}{\Sigma x_2^2 (1 - r_{12}^2)} \\
\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) &= \frac{- r_{12} \cdot \hat{\sigma}^2}{(1 - r_{12}^2)\sqrt{\Sigma x_1^2 \ \Sigma x_2^2}}
\end{aligned}
\right\} \quad \ldots \quad (2-7)
$$

وعملياً فان المقدرات تمتلك تبايناً وتبايناً مشتركاً كبيراً اعتماداً على العلاقات في (2-7).

– وهذا بدوره يقود الى ان اختبار (t) لاحد المعلمات (أو أكثر) يكون غير معنوي في حين معامل التحديد R^2 يكون عالياً.

– كما ان مجال الثقة للمعلمات المقدرة يكون كبيراً مما يدفع الباحث الى قبول فرضية العدم بشكل اكبر.

– وفوق ذلك كله فان المقدرات وانحرافها المعياري ربما يكون حساساً للتغيرات في البيانات. فقد يتولد تغيير ملموس عند تغيير بسيط في حجم العينة.

بمعنى عند توسيع حجم العينة بمقدار بسيط (كأن يزداد عدد المشاهدات بمقدار مشاهدتين) ينتج مقدرات مختلفة تماماً وربما يكون الاختلاف حتى في اتجاه العلاقة (قيمة المعلمات سالبة بعد ان كانت موجبة).

(5-7)عامل تضخيم التباين (VIF) (Variance inflation factor)

في نموذج الانحدار الذي يتضمن متغيرين توضيحيين X_1 و X_2 :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

تعرف النسبة $\left(\frac{1}{1 - r_{12}^2}\right)$ بمعامل تضخيم التباين.

اذ يتضح من العلاقة (2-7) ان ازدياد معامل الارتباط بين المتغيرات التوضيحية(r_{12}) يؤدي الى تضخيم التباين للمعلمات المقدرة بمقدار $\left(\frac{1}{1 - r_{12}^2}\right)$. ويرمز له بـ (VIF) .

بعبارة أخرى فأن:

$$var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\Sigma x_1^2} \cdot (VIF)$$

$$var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\Sigma x_2^2} \cdot (VIF)$$

وبشكل عام في نموذج الانحدار:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \quad (3-7) \quad ...$$

فان (VIF) يعتمد على معامل الارتباط المساعد (Auxiliary correlation) (R_j^2) ويعرف على وفق المعادلة :

$$VIF = \frac{1}{1 - R_j^2} \quad ... \quad (4-7)$$

حيث ان (R_j^2) يمثل معامل الارتباط المتعدد بين المتغير X_j والمتغيرات التوضيحية الأخرى، ويمكن حسابها من علاقة الانحدار X_j مع باقي المتغيرات التوضيحية.

$$X_j = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \cdots + \alpha_k X_k + u \quad (5-7) \quad ...$$

ثم نحسب معامل التحديد للعلاقة (5-7) وهو ما يسمى R_j^2 .

مثال(7-1): العلاقة بين معامل تضخيم تباين المقدرات وبين معامل الارتباط البسيط بين المتغيرين X_1 و X_2 في علاقة الانحدار المتعدد بمتغيرين توضيحيين يمكن توضيحها من خلال الجدول التالي:

يبين الجدول ان معامل تضخيم التباين في تزايد بشكل متسارع مع معامل الارتباط فكلما ازدادت قيمته عن (2) تسارعت قيمة التباين بالتزايد. وهذا يشير الى ان ارتفاع قيمة معامل الارتباط (r_{12}) عن (0.7) فأن قيمة تباين المعلمة المقدرة يتسارع بالتزايد وبشكل مضطرد.

الجدول (٧-١)

علاقة معامل الارتباط بتباين معلمة الانحدار

r_{12}	0	0.5	0.7	0.8	0.9	0.95	0.99	0.999
VIF	1	1.33	1.96	2.78	5.76	10.26	50.25	500.00
$var(\hat{\beta}_1)$	$\frac{\hat{\sigma}^2}{\Sigma x_1^2} = A$	1.33A	1.96A	2.78A	6.76A	10.26A	50.25A	500.0A

(6-7)الكشف عن وجود مشكلة التعدد الخطي Detecting multicollinearity

ذكر الباحث (Kmenta) تحذيراً مهماً يمكن تلخيص ملامحه كآلاتي:

1- مشكلة التعدد الخطي مسألة تتعلق بالدرجة وليست بالنوع. إذ غالباً ما تكون هناك علاقة بين المتغيرات المستقلة نظراً للتأثير الحاصل بين بعضها ببعض. وفي حال انعدام العلاقة بين المتغيرات المستقلة (التوضيحية) انتفت الحاجة الى استخدام نموذج انحدار خطي متعدد، ويستعاض عنه بـ (k) من النماذج الخطية البسيطة والذي يحتوي كل واحد منها على العلاقة بين المتغير المعتمد وواحد من المتغيرات المستقلة.

2- مشكلة التعدد الخطي هي مشكلة عينة وليست مشكلة مجتمع، فأساس المشكلة يعود الى الافتراض بان المتغيرات التوضيحية هي متغيرات غير عشوائية أي ثابتة للعينة المختارة.

واستناداً الى ذلك التحذير فأننا لا نختبر وجود أو عدم وجود المشكلة وإنما نقيس درجتها في العينة. أضف إلى ذلك أن المشكلة تظهر في الغالب بسبب جمع البيانات في اغلب العلوم الاجتماعية لذلك توجد طرائق متعددة للكشف عن المشكلة بعضها بسيط Rule of thumb والآخر إحصائي Formal أو informal.

(1-6-7) مؤشرات وجود التعدد الخطي : (Rule of thumb)

نتطرق الى أهم المؤشرات التي تلازم التعدد الخطي المؤثر فى النتائج:

1- عندما يكون معامل التحديد للنموذج (R^2) عالياً وكذلك(الاحصاءة F) للنموذج الكلي تكون معنوية ، في حين تكون قيم(الاحصاءة t) لبعض المعلمات المقدرة غير معنوية. وتعد هذه سمة ملازمة لمشكلة التعدد الخطي.

2- دراسة معاملات الارتباط بين المتغيرات التوضيحية (سواء كان البسيط أم الجزئي) مهمة وذات فائدة غير أنها أدلة ليست كافية لتأشير درجة التعدد الخطي. فهي ليست شروطاً ضرورية على الرغم من انها شروط كافية. فالتعدد الخطي قد يكون مؤثراً فى النتائج بالرغم من صغر معاملات الارتباط.

3- معيار F الجزئي للانحدار المساعد، اذ تشير معنويته الى ارتباط احد المتغيرات التوضيحية بالمتغيرات التوضيحية الأخرى. ويمكن حسابه (بالنسبة للمتغير التوضيحي X_1) على وفق الآتي:

$$F_1^* = \frac{(\mathrm{R}^2_{X_1 \cdot X_2 X_3 \cdots X_k})/(k-1)}{(1-\mathrm{R}^2_{X_1 \cdot X_2 X_3 \cdots X_k})/(\mathrm{n}-\mathrm{k})} \sim F_{(k-1, n-k, 1-\alpha)}$$

حيث ان $(\mathrm{R}^2_{X_1 \cdot X_2 X_3 \cdots X_k})$ تمثل معامل التحديد لانحدار X_1 ضد X_2 ، X_3 ، ... ، X_k. وأختصاراً يرمز له ب (R_1^2) وتكون العلاقة كالآتي:

$$F_1^* = \frac{R_1^2 / k-1}{(1-R_1^2)/(n-k)}$$

وبشكل عام للمتغير التوضيحي X_j :

$$F_j^* = \frac{R_j^2 / k-1}{(1-R_j^2)/(n-k)} \qquad \forall \quad j=1,\cdots,k \qquad \cdots \qquad (6-7)$$

4- اختبار (klein) تشير قاعدة الابهام (لكلا ين) بان التعدد الخطي يكون مشكلة مؤثرة في النتائج وتتطلب مداخلة للعلاج اذا (R_j^2) للانحدار المساعد يكون اكبر من (R^2) للنموذج الكلي، $R_j^2 \rangle\ R^2$.

5- معامل تضخيم التباين (VIF) يعد تشخيصاً إضافيا، فمع ازدياد قيمة VIF عن (2) يكون دليلاً مهماً على وجود المشكلة بشكل يؤثر فى النتائج. بالرغم من انه لا يعد شرطاً ضرورياً ولا كافياً.

ولاختبار مشكلة التعدد الخطي باستخدام اختبار (klein) لبيانات المثال(٥ – 14) نفسها واعتمادا على البرنامج الجاهز (SPSS)،ندرس علاقة ارتباط بيرسن بين المتغيرين X_1 و X_2 والتي يحسبها البرنامج:

$r_{X1X2} = 0.332$ ثم نربعها $(r^2_{X_1X_2} = 0.11022)$ وتقارن مع معامل التحديد للنموذج ($R^2 = 0.646$) وهذا دليل على خلو النموذج من مشكلة التعدد الخطي على وفق اختبار(klein)،

(2-6-7) الاختبارات الاحصائية: " Formal Statistical – test "

1- اختبار " Beaton – Glauber "

تكون خطوات الاختبار بان نحسب مصفوفة معاملات الارتباط البسيط بين المتغيرات التوضيحية والتي يرمز لها ب (R)، ثم نحسب معكوس هذه المصفوفة(R^{-1}) ونتفحص العناصر القطرية لهذه المصفوفة والتي تحسب على وفق الصيغة:

$$r^{kk} = \frac{|\ R_{kk}|}{|\ R\ |} \quad \cdot\ \cdot\ \cdot \qquad (7-7)$$

فعندما تكون قيمة العناصر القطرية قريبة من الواحد فهذا يؤشر عدم وجود مشكلة التعدد الخطي، في حين ابتعاد هذه القيمة عن الواحد (كبرها) يشير الى ان النتائج تتأثر بمشكلة التعدد الخطي .

وبعبارة اخرى:

اذا $r^{kk} = 1$ فان المتغيرات التوضيحية متعامدة .

$r^{kk} \to \infty$، ان المتغير X_k ينذر بخطر التعدد الخطي.

2– اختبار فاراركليبر " Farrar - Glauber-test "

هذا الاختبار يجري على ثلاث مراحل.

المرحلة الأولى: يستخدم اختبار مربع كاي للكشف عن درجة (شدة) التداخل الخطي المتعدد وتكون

فرضية العدم : المتغيرات التوضيحية متعامدة H_0: $X's$ orthogonal

مقابل الفرضية البديلة : المتغيرات التوضيحية مترابطة $X's$ H_1 : multicollinear

والمختبر الذي يستخدم هو :

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2k+5) \right] Ln \mid R \mid \quad \ldots \quad (8-7)$$

حيث ان | R | يمثل محدد مصفوفة معاملات الارتباط البسيط .

Ln تمثل اللوغاريتم للأساس الطبيعي .

k : عدد المتغيرات المستقلة التوضيحية .

n: حجم العينة المستخدمة .

وتقارن القيمة المحسوبة لمربع كاي مع القيمة الجدولية بدرجة حرية $\frac{k(k-1)}{2}$ ومستوى دلالة% α.

ومع رفض فرضية العدم يجب الانتقال الى المرحلة الثانية .

المرحلة الثانية: تحديد موقع التعدد الخطي "Location of multicollinearity"

يتم حساب معامل الارتباط المتعدد لكل علاقة انحدار مساعد، أي نحسب R_j^2 (j =1,...k)

لعلاقة انحدار X_j كمتغير معتمد مع باقي المتغيرات التوضيحية والمقطع الثابت، ويحسب معامل التحديد لكل علاقة انحدار مساعد.

وتختبر الفرضية :

$$H_0: R_j^2 = 0$$
$$vs. \quad H_1: R_j^2 \neq 0 \qquad j = 1,\ldots,k$$

وباستخدام المختبر الإحصائي في العلاقة (7-6) :

$$F_j^* = \frac{R_j^2/(k-1)}{(1-R_j^2)/(n-k)} \sim F_{(k-1,n-k,1-\alpha)} \qquad , \forall \quad j = 1,\cdots,k$$

ورفض فرضية العدم تشير الى ان المتغير X_j ذو تداخل خطي متعدد مع المتغيرات التوضيحية الأخرى. وذلك يقترح الانتقال الى المرحلة الثالثة.

المرحلة الثالثة : تحديد نمط التعدد الخطيPattern" "of multicollinearity"

وتتطلب هذه المرحلة الخطوات التالية .

حساب معاملات الارتباط الجزئي بين المتغيرات التوضيحية.

نختبر معنويتها الإحصائية باعتماد اختبار t الجزئي.

لفرضية العدم $H_0: r_{xj,xi.\ others\ constant} = 0$

vs.: $H_1: r_{xj,xi\ .all\ others\ constant} \neq 0$

والمختبر المستخدم هو:

$$t_{ij.\ others\ constant} = r_{ij.\ others\ constant} \cdot \sqrt{\frac{n-k}{1-r_{ij.\ others\ constant}}} \qquad \cdots \qquad (9-7)$$

علماً ان معامل الارتباط الجزئي يحسب على وفق القانون:

$$r_{ij.l} = \frac{r_{ij} - r_{il} r_{jl}}{\sqrt{1-r_{il}^2}\sqrt{1-r_{jl}^2}} \quad , \quad \forall \quad i, j, l = 1, \cdots, k$$

حيث ان r_{ij} تمثل معامل الارتباط البسيط بين المتغيرين X_i و X_j ، ومع رفض فرضية العدم فان القرار يكون بان المتغير X_i هو المسؤول عن وجود التداخل الخطي المتعدد مع المتغير X_j الذي أفرزته المرحلة الثانية.

(7-7) طرائق التخفيف من حدة المشكلة.Remedies measures

1- ليس من السهل ايجاد حلول ناجعة لمشكلة التعدد الخطي بل يرى بعضهم بترك النتائج كما هي حيث ان أية معالجة لتخفيف حدة المشكلة لها مسأوئ من نوع أو آخر وبذلك ترك النتائج كما هي هو الاسلوب الاصوب. ووجهة النظر هذه ترى ان وجود التعدد الخطي لا يعمل دائماً على تخفيض نسبة t وجعلها غير معنوية أو تعمل على تغيير قيمة المعلمات بجعلها مناقضة لما يتوقعه الباحث حولها. واستخدام طرائق التخفيف فقط في حال كون المشكلة تؤثر في النتائج. فقد نجد معامل ارتباط بين بعض المتغيرات التوضيحية يقارب (0.97) وتبقى معلمات هذه المتغيرات معنوية بموجب النسبة t وبذلك ليس بالضرورة عمل شيء من اجل التلطيف لان أي معالجة ربما تولد مشكلة اخرى للنموذج، فمشكلة التعدد الخطي في بعض النماذج مماثله لبعض الامراض غير المهددة لحياة الانسان، فمخاطرة اجراء عمليه تكون فقط اذا كان المرض يحدث مشكلة مهمة. ومن جانب اخر فان حذف متغير من معادلة يكون في غاية الخطر لانه سوف يحدث تحييزاً للتوصيف، ولذا في الغالب تترك مشكلة التعدد الخطي بالرغم من انخفاض (قيمة t).

2- حذف احد المتغيرات التوضيحية ذات علاقة ارتباط عالية. .وتراعى ملاحظة سبب الارتباط فقط يكون احد المتغيرات زائداً أو أن احد المتغيرات هو الذي يفضل ان يضمن في المعادلة فمثلاً عدد السكان

والدخل المتاح يقيسا الشيء نفسه وهو حجم السوق في دراسة الطلب لذا يفضل إضافة احدهما فقط في معادلة الطلب. وبذلك يعتمد على هدف الباحث وفرضيته.

3- تحويل المتغيرات التي تعاني من ارتفاع علاقة الارتباط بينهما إذا كان كلاهما مهماً في دراسة الباحث ونوع التحويل يكون :

أ) خلق تركيب بدلالة كلا المتغيرين وتضمينه كمتغير جديد في المعادلة وخاصة إذا كان الهدف لإغراض التنبؤ . أو استخدام القيود المسبقة من دراسات تطبيقية سابقة أو بالاعتماد على النظريات في اطار معين، على سبيل المثال استخدام دالة الانتاج من نوع كوب دوجلاص تفترض عائد حجم ثابت، لذا يتم استخدام هذا القيد في المعادلة يعمل على تحويل صيغة المتغيرات الى النسب (الناتج/العمل ، رأس المال / العمل) وتسمى هذه الطريقة المربعات الصغرى المقيدة.

ب) تحويل المتغيرات باستخدام الفرق الأول. وحيث ان التعدد الخطي غالباً ما يوجد في دراسات السلاسل الزمنية فان استخدام الفرق الأول يعمل على تحريك التغيرات إلى الأعلى بنسبة اقل من القيم الأصلية. وجدير بالذكر ان استخدام الفرق الأول لا يصلح لدراسات المقاطع العرضية. كما ان استخدامه في دراسات السلاسل الزمنية يخلق مشكلة الارتباط الذاتي فضلا عن انه يعمل على فقدان درجات الحرية.

ج) كما ان استخدام القيم المعيارية للمتغيرات التوضيحية بدلاً من القيم العادية يكون ملائماً للتخفيف من حده التعدد الخطي علماً بان القيم المعيارية هي :

$$X_i^* = \frac{X_i - \overline{X}}{s.e(X_i)} \qquad \forall \quad i = 1, \cdots, n$$

4- العمل على زيادة حجم العينة. وإحدى أهم طرائق زيادة حجم العينة هو دمج المقاطع العرضية مع السلاسل الزمنية . حيث أن تضمين المقاطع العرضية التي تكون غير متداخلة مع السلاسل الزمنية تعمل على تقليص التعدد الخطي في النموذج الكلي[*]. وتبقى المشكلة مرتبطة بتفسير النتائج وطريقة التقدير. وبعضهم يرى مشكلات تفسير النتائج في حالة الدمج تكون أسوأ من التبعات التي تخلقها مشكلة التعدد الخطي .

5- هناك بعض الطرق التي تعالج المشكلة إحصائياً مثل : التحليل العاملي(Factor Analysis)، المركبات الرئيسة (principle component)، طريقة الحرف(Ridge Regression) أو طريقة

(*) مثلاً لدراسة الطلب على سلعة في قطر معين اقترح الاقتصادي (Tobin) استخدام بيانات مقطعية من ميزانية الأسرة للحصول على المرونة الدخلية ثم استخدامها في دالة الطلب للسلاسل الزمنية للحصول على المرونة السعرية.

الانحدار المتسلسل (Stepwise Regression) أو طريقة التقدير المختلط المقترحة من قبل (ثايل وكولد بيرجر).

وجدير بالذكر إن لكل طريقة من طرائق التخفيف من حدة المشكلة مساوئ وأضرار جانبية منها ما يخلق مشكلات أخرى تحتاج إلى حل وتضع الباحث في دوامة مستمرة، وبعضهم الآخر يخلق مشكلات تتعلق بتفسير نتائج التقدير. وتبقى خبرة الباحث وهدف الدراسة هي الفيصل في اختيار طريقة دون سواها أو ترك النتائج دون تعديل. علماً بان التعدد الخطي لا يشكل خطراً مهماً اذا كان الهدف هو التنبؤ فالمعيار في هذه الحالة هو كبر R^2 على شرط ان المتغيرات التوضيحية تتبع نمط الارتباط نفسه في فترة التنبؤ كما في فترة البيانات المستخدمة في الدراسة.

أسئلة الفصل السابع:

س١: أ) وضح مفهوم التعدد الخطي وانواعه.

ب) اذكر أهم مصادر ظهور مشكلة التعدد الخطي.

ج) اذكر أهم الملامح التي تشير الى وجود مشكلة التعدد الخطي.

س2: صحح العبارات الخاطئة ان وجدت.

١. وجود المشاهدات الشاذة يتسبب بمشكلة التعدد الخطي.
٢. معامل تضخم التباين هو مقلوب معامل التحديد.
٣. أحد الإشارات لمشكلة التعدد الخطي ان تكون قيمة(R^2) عالية في حين (F) للنموذج ككل عالية و (t) ايضاً عالية.
٤. (VIF) شرط ضروري وكافي للحصول على تباين كبير.
٥. تباين المعلمات المقدرة في النموذج المتعدد يرتبط عكسياً بمعامل التحديد.
٦. وجود التعدد الخطي التام يولد مقدرات متحيزة وغير كفوءة.
٧. اختبار فارار كليبر يعتمد على الاحصاءة χ^2 بدرجات حرية (k) .
٨. طريقة المركبات الأساسية تخلص النموذج تماماً من مشكلة التعدد الخطي، ولذلك فهي طريقة مرغوبة.
٩. وجود مشكلة التعدد الخطي يمكن التعايش معها اذا كان الهدف الرئيس هو تقدير معلمات النموذج.
١٠. رتبة مصفوفة X تامة من ناحية السطور تعني وجود تغيرات واضحة وملموسة لمشاهدات المتغيرات التوضيحية.

س3: حدد أهم طرائق الكشف عن مشكلة التعدد الخطي.

س4: في النموذج $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$ وضح هل ان النموذج يعاني من مشكلة التعدد الخطي.

س5: اشرح أثر الارتباط المتعدد في حالة كون قيم أحد المتغيرات التوضيحية متسأوية المشاهدات.

س6: أعط مثالاً توضح فيه إن معاملات الارتباط البسيط بين المتغيرات التوضيحية لا تكون كافية للكشف عن وجود مشكلة التعدد الخطي.

س7: اذا علمت ان مصفوفة الارتباط البسيط بين المتغيرات التوضيحية لنموذج مكون من ثلاث متغيرات توضيحية ولعينة مكونة من 60 مشاهدة هي:

$$R = \begin{pmatrix} 1 & 0.879 & -0.339 \\ & 1 & -0.305 \\ & & 1 \end{pmatrix}$$

هل ان النتائج تتأثر بمشكلة التعدد الخطي باستخدام:

أ) اختبار Beaton&Glauber .

ب) اختبار Farrar – Glauber .

الفصل الثامن
الانحدار غير الخطي Nonlinear regression

تمهيد

أن أول فرضيات التحليل في الانحدار هي **الفرضية (1)** هذه الفرضية تنص على ان علاقة الانحدار تكون خطية بدلالة المعلمات.

ومعلوم ان العلاقات أما تكون خطية بدلالة المعلمات وبدلالة المتغيرات وهي العلاقة التي تمت دراستها في الفصول السابقة من الكتاب.

العلاقة العامة التي تتضمن (k) من المتغيرات التوضيحية والتي يمكن كتابتها بالصيغة:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t$$

أو

$$Y_t = \sum_{j=0}^{k} \beta_{jt} X_{jt} + u_t \qquad \forall \qquad t = 1,\ldots,n$$

$$j = 1,\cdots,k$$

هي علاقة خطية بدلالة المعلمات، وكذلك بدلالة المتغيرات.

أو ان تكون خطية بدلالة المعلمات وغير خطية بدلالة المتغيرات (ومعادلاتها تسمى علاقات الانحدار غير الخطية) والتي ستتم دراستها في هذا الفصل والنوع الثالث ان تكون علاقة الانحدار غير خطية بدلالة المعلمات ومن أمثلتها:

$$Y = \beta_o + \sqrt{\beta_1} X + u \qquad \ldots \qquad (1)$$

$$Y = \beta_o + \beta_1^{\,2} X + u \qquad \ldots \qquad (2)$$

$$Y = \ln\left(\beta_o + \beta_1 X\right) + u \qquad \ldots \qquad (3)$$

وهكذا . . .

فالمعادلات (١) و(٢) هي غير خطية بدلالة المعلمات ولكنها خطية بدلالة المتغيرات في حين العلاقة (٣) غير خطية بدلالة المعلمات والمتغيرات.

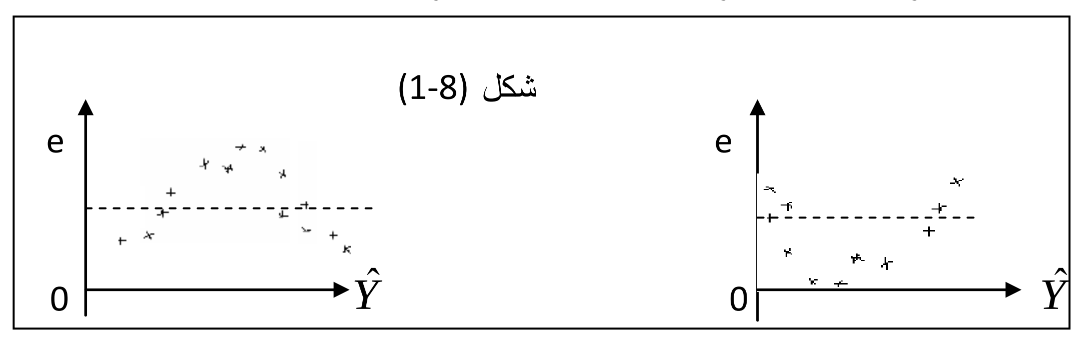
وهذا النوع من علاقات الانحدار لا يمكن تقدير معلماته باستخدام معيار المربعات الصغرى الاعتيادية بل يتم استخدام اسلوب الانحدار غير الخطي المتكرر، والذي هو خارج مفردات هذا الكتاب.

ولقد أوضحنا في الفصول السابقة من هذا الكتاب بان اختبار العلاقة الخطية يمكن ان يعبر عنه بأختبار F للنموذج الكلي والذي يختبر الفرضية

$$H_o = \beta_1 = \beta_2 = \ \ldots \ = \beta_k = 0$$

وذلك باستخدام جدول تحليل التباين.

ويمكن عمل الاختبار بشكل أولي من خلال رسم الانتشار في حالة النموذج البسيط أو باستخدام تحليل البواقي وذلك برسم الأخطاء أو الأخطاء القياسية أو مربع الأخطاء على المحور العمودي بدلالة $\hat{Y}_i$ أو X_iعلى المحور الأفقي. فإذا كان الشكل الناتج يوضح نمطاً معيناً كما في الشكل(٨−١) فذلك يدلل على ان العلاقة غير خطية أو ان فرضية العلاقة الخطية غير متحققة.



شكل (1-8)

مثال(8−١): اذا توفرت البيانات للمتغير المعتمد Y والمتغير المستقل X فهل ان معادلة الخط المستقيم تلائم البيانات.

$$Y' = [1 \quad 4 \quad 3 \quad 8 \quad 9]$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix} \quad , \quad (X'X)^{-1} = \begin{bmatrix} 0.6 & -0.2 \\ & 0.1 \end{bmatrix} \quad , \quad X'Y = \begin{bmatrix} 25 \\ 70 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

معادلة الانحدار المقدرة: $\hat{Y} = 1 + 2X$

ولمعرفة مدى ملاءمة البيانات مع علاقة الخط المستقيم، يتم اعتماد طريقة الرسم أولاً والتي يمكن انجازها اما باستخدام رسم البواقي والذي يتم برسم قيم X_i أو $\hat{Y}_i$ على المحور الافقي وقيم الأخطاء على المحور العمودي. أو يمكن اعتماد رسم الانتشار للتأكد من تحقق ذلك.

ولاستخدام تحليل البواقي

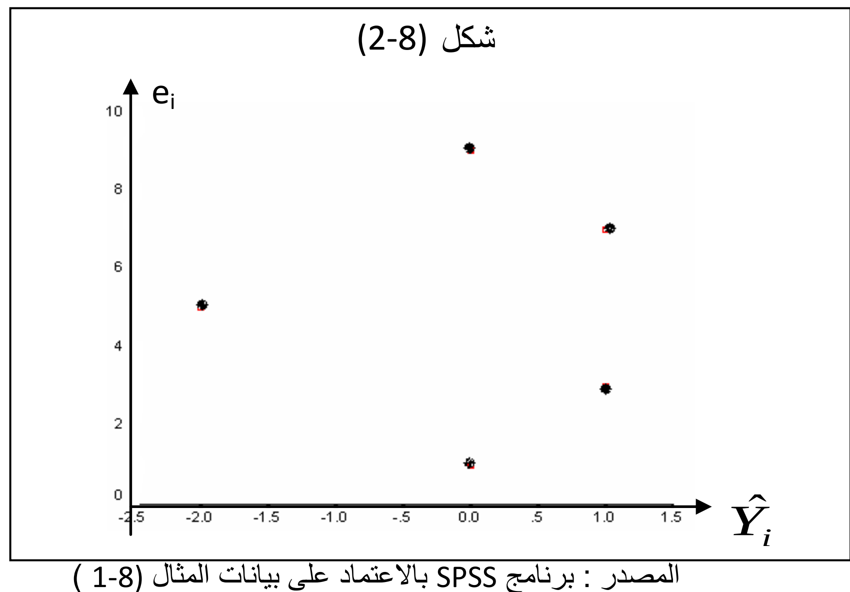
تحسب القيم $\hat{Y}_i$ و e_i لكل قيمة من قيم X

$$\hat{Y}' = [1 \quad 3 \quad 5 \quad 7 \quad 9] \quad , \quad e' = [0 \quad 1 \quad -2 \quad 1 \quad 0]$$

$$\Sigma e^2 = RSS = 6$$

ثم نرسم e_i ضد $\hat{Y}$ فيتضح الشكل (8–2)

شكل (2-8)

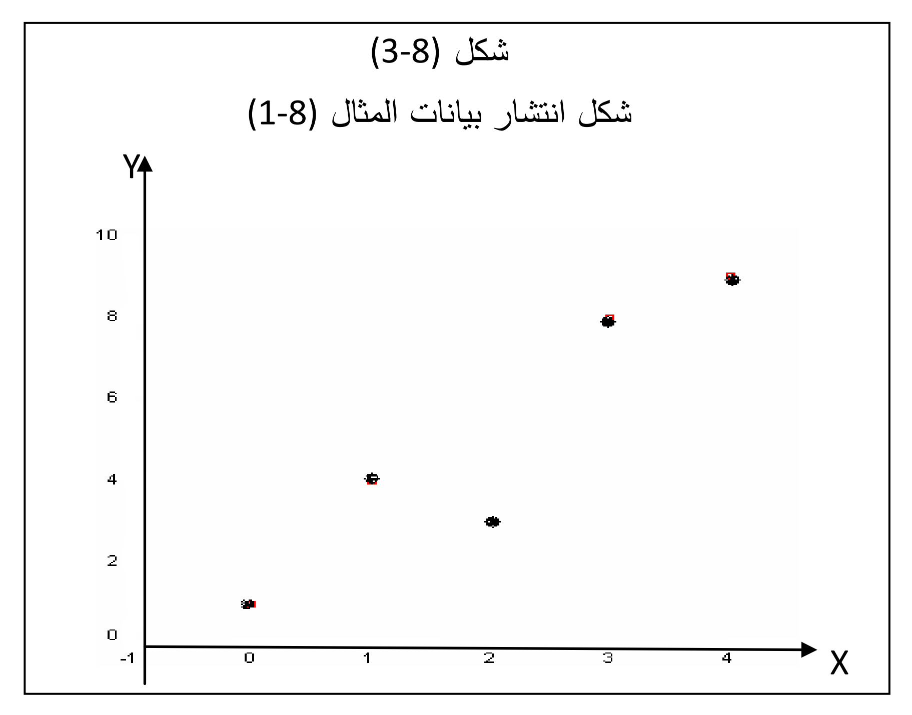


المصدر : برنامج SPSS بالاعتماد على بيانات المثال (8-1)

اما بالاعتماد على رسم الانتشار فيتم وضع قيم Y على المحور العمودي وقيم X على المحور الافقي كما في الشكل (٨–٣)

شكل (3-8)

شكل انتشار بيانات المثال (8-1)



المصدر: برنامج SPSS على بيانات المثال (8-1)

نؤكد في هذا الصدد ان طريقة الرسم تعتمد على الحكم الشخصي وخاصة مع صغر حجم العينة وكذلك تعتمد على خبرة الباحث، لذا لابد من اقترانها بالاختبارات الإحصائية النظامية ومنها جدول تحليل التباين.

$$H_o :\ \beta_1 = 0 \quad vs \quad H_1 :\ \beta_1 \neq 0$$

وباستخدام جدول تحليل التباين

جدول (٨-٣)

جدول تحليل التباين للمثال (٨-١)

S.O.V	df	SS	MS	F	
Regression	1	40	40	20	$F_{c(1,3,0.95)} = 10.13$
Error	3	٦	2		

وبمقارنتها مع القيمة الجدولية يتضح ان العلاقة الخطية ملائمة.

(٨-١) <u>خطية العلاقة بدلالة المعلمات:</u>

لقد تم التأكيد في الفصول السابقة على ان النماذج التي تعتمد في الانحدار هي خطية بدلالة المعلمات، وقد تكون هذه النماذج خطية (أو غير خطية) بدلالة المتغيرات. وفي الواقع العملي هناك العديد من العلاقات التي تكون غير خطية بدلالة المتغيرات ولكنها خطية بدلالة المعلمات ومنها:

نموذج اللوغاريتم – الخطي Log - liner model

نموذج شبه اللوغاريتم semi log model

نموذج المعكوس Reciprocal model

نموذج متعدد الحدود polynomial model

وغيرها العديد.

ولاختيار الأشكال الدالية المختلفة لابد من إعطاء عناية خاصة لأمرين مهمين هما حد متغير الاضطراب العشوائيu_i ، اما الأمر الاخر فهو يخص تفسير المعلمات.

(1-1-8) نموذج اللوغاريتم الخطي Log linear model:

لنمعن النظر في العلاقة: $Y_i = \beta_1 X_i^{\beta_2}$

فهي علاقة أسية "exponential" (غير خطية بدلالة β) ولإغراض التقدير يمكن إعادة صياغتها كالآتي:

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \quad \cdots \quad (1-8)$$

أو $$Y_i = \beta_1 X_i^{\beta_2} u_i \quad \cdots \quad (2-8)$$

أو $$Y_i = \beta_1 X_i^{\beta_2} + u_i \quad \cdots \quad (3-8)$$

وبأخذ لوغاريتم الطرفين تعاد كتابتها كالآتي:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + \ln e^{u_i} = \alpha + \beta_2 \ln X_i + u_i \quad \cdots \quad (a1-8)$$

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + \ln u_i = \alpha + \beta_2 \ln X_i + \ln u_i \quad \cdots \quad (a2-8)$$

$$\ln Y_i = \ln(\beta_1 \cdot X_i^{\beta_2} + u_i) \quad \cdots \quad (a3-8)$$

حيث ان $\alpha = \ln \beta_1$

فالعلاقة (1-8) خطية بدلالة المعلمات اذ بأخذ اللوغاريتم للطرفيين تم تحويله الى خطي بدلالة المعلمات α و β_2 وكذلك الحال بالنسبة للعلاقة (2-8). غير انه لا بد من التحقيق حول صفات المتغير العشوائي في المعادلات (a1-8) و (a2-8). فمع افتراض ان المتغير العشوائي u_i يتوزع طبيعياً. بمتوسط صفر وتباين متجانس وارتباط ذاتي صفري . $u_i \sim N(0,\sigma^2)$

فالمتغير العشوائي في المعادلة(a٢−8) ، ($\ln u_i$) ، لا يكون كذلك. وعليه لابد من إعطاء عناية خاصة لحد الخطأ عند تحويل النموذج لأغراض الانحدار.

أما العلاقة (a3-8) فهي في حقيقة الأمر غير خطية بدلالة المعلمات. ولا توجد طريقة بسيطة لأخذ لوغاريتم الطرفين لهذه العلاقة ، وعليه يجب حله باستخدام طرائق غير خطية.

اما بخصوص معلمات العلاقة (1-8) والتي تعد خطية بدلالة لوغاريتم المتغيرات Y و X عند اخذ لوغاريتم الطرفين كما في العلاقة (a1-8) والتي يمكن تقديرها باستخدام OLS فتكون المقدرات $\hat{\alpha}$ و $(\hat{\beta}_2)$ مقدرات غير متحيزة وخطية وباقل تباين للمعلمات α و β على التوالي:

علماً بان $\hat{\beta}_1 = e^{\hat{\alpha}}$ وبذلك فان $\hat{\alpha} = (\ln \hat{\beta}_1)$ حيث ان: $(e = 2.718)$

اما $\hat{\beta}_2$ فهي تقيس مرونة Y بالنسة لـ X لان

$$\hat{\beta}_2 = \frac{d \ln Y}{d \ln X} = \frac{\frac{dY}{Y}}{\frac{dX}{X}}$$

$$= \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{\text{القيمة الحدية}}{\text{القيمة المتوسطة}}$$

وحيث ان $\hat{\beta}_2$ تكون ثابتة لذا يسمى النموذج نموذج المرونة الثابتة (Constant elasticity model)

(2–1–8) نماذج شبة اللوغاريتم Semi–Logarithm: ويمكن ان تكون اما :

أ. شبه لوغاريتمية بدلالة Y " Log – Lin Model "

ب. شبه لوغاريتمية بدلالة X " Lin – Log Model "

أ. شبه لوغاريتمية بدلالة Y: " Log – Lin Model "

ان المعادلة التي تستخدم لاستخراج معدلات النمو المركب هي :

$$Y_t = Y_o(1+r)^t \quad . \; . \; . \quad (4\text{-}8)$$

حيث Y_0 تمثل القيمة الأولية (الابتدائية) للمتغير Y .

r : معدل النمو المركب

t : السنوات للسلسلة Y .

وهي معادلة غير خطية بدلالة Y_0 .

وبأخذ لوغاريتم الطرفين تتحول الى خطية

$$\ln Y = \ln Y_o + t(\ln(1+r)) \quad . \; . \; . \quad (5\text{-}8)$$

$$\ln Y = \beta_1 + \beta_2 \; t$$

حيث ان :

$$\beta_1 = \ln Y_o \quad . \; . \; . \quad (6\text{-}8)$$

$$\beta_2 = \ln(1+r) \quad . \; . \; . \quad (7\text{-}8)$$

وبإضافة المتغير العشوائي u تصبح العلاقة

$$\ln Y = \beta_1 + \beta_2 t + u \quad . \; . \; . \quad (8\text{-}8)$$

العلاقة (8–٥) تسمى شبه لوغارتمية (Semi log model) بدلالة Y وتسمى (Log-Lin model) أيضاً.

– ولإغراض التقدير يمكن إعادة المعادلة (8-4) بالصيغ التالية:

$$Y_t = Y_o(1+r)^t * u \qquad \cdots \quad (9-8)$$

أو

$$Y_t = Y_o(1+r)^t * e^u \qquad \cdots \quad (10-8)$$

أو

$$Y_t = Y_o(1+r)^t + u \qquad \cdots \quad (11-8)$$

ويتبين كما في الفقرة (٨–١–١) ان (8-9) و(8-10) تصبح خطية بدلالة المعلمات مع التحفظ حول توزيع حد الاضطراب اما (8-11) فهي غير خطية بدلالة المعلمات.

تفسير المعلمات:

المعلمة $(\ln(1+r) = \beta_2)$ تقيس التغير النسبي في Y نسبة الى تغير مطلق بقيمة المتغير t.

$\beta_2 = \frac{d \ln y}{dt}$ وهي بذلك تقيس معدل النمو الجاري (instantaneous). وبذلك فالعلاقة (8-8) تسمى نموذج النمو الثابت constant growth model.

اما معدل النمو المركب فيمكن حسابه من العلاقة (8–7) : $\left(r = (e^{\beta_2} - 1)100\ \%\right)$

المعلمة β_1 تمثل المقطع الثابت وبموجب العلاقة (٨–٦) فان قيمة التغير المعتمد في بداية المدة يمكن حسابه: $\left(Y_0 = e^{\beta_1}\right)$

ملاحظة: ان نتائج تقدير العلاقة (٨–٨) تعتمد على استقرارية السلسلة.

ب. نموذج شبه اللوغارتمي بدلالة X.(Lin-Log Model)

صيغة النموذج : $Y = \beta_1 + \beta_2 \ln X + u$

المعلمة β_2 تمثل التغير المطلق في Y نسبة الى التغير النسبي بـ X كالآتي:

$$\beta_2 = \frac{\Delta Y}{\frac{\Delta X}{X}}$$

فمع تغير $\frac{\Delta X}{X}$ بمقدار وحدة واحدة فان تغيرا مطلقا في Y يكون بمقدار $((0.01)\beta_2)$ اما القيمة الحدية أو الأثر للمتغير X على(Y) فيمكن حسابه على وفق العلاقة:

$$\frac{dY}{dX} = \frac{\beta_2}{X}$$

(٨-١-٣) **نموذج المعكوس (Reciprocal Model)**

الصيغة العامة للنموذج:

$$Y = \beta_1 + \beta_2 \frac{1}{X} + u \qquad . . . \qquad (12\text{-}8)$$

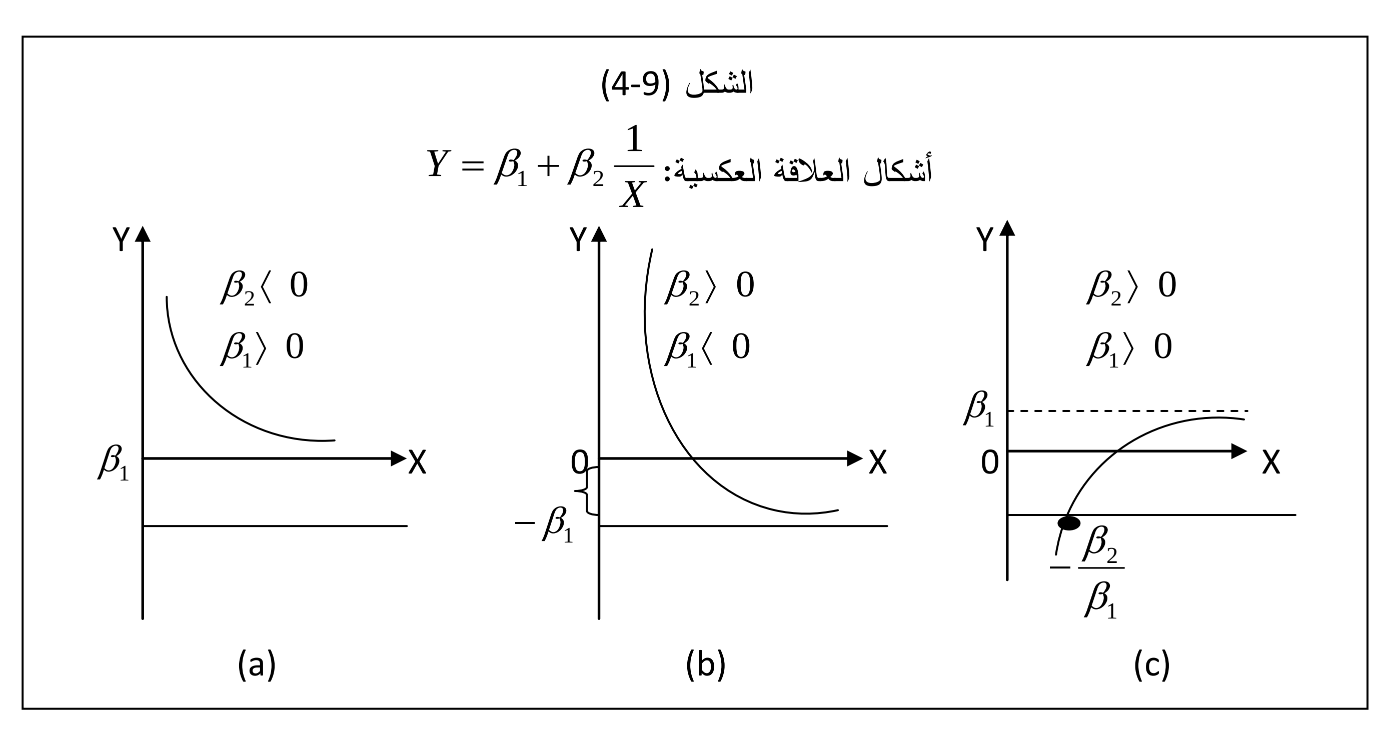
وبالرغم من ان العلاقة (12-8) هي غير خطية بدلالة X ولكنها خطية بدلالة المعلمات β_1 و β_2

وذلك من خلال افتراض $X^* = \frac{1}{X}$ فتتحول الصيغة إلى:

$$Y = \beta_1 + \beta_2 X^* + u \qquad . . . \qquad (a12\text{–}8)$$

ويتميز النموذج (12–8) بالميزات التالية:

مع زيادة X بشكل لا نهائي فان $(\beta_2 \frac{1}{X})$ تنخفض الى الصفر وبذلك فان Y تبلغ β_1 والتي تسمى المحاذي (Asymptote) أو الغاية (Limit) والشكل (8–4) يوضح بعض اشكال المنحنيات التي تتمثل بالعلاقة (٨-١٢)



وعند تقدير العلاقة (a12-8) فان تفسير معلمات الانحدار β_2 لا تمثل القيمة الحدية بل ان اثر المتغير X على المتغير Y يمكن حسابها كالآتي:

$$\frac{dY}{dX} = -\beta_2 \cdot \frac{1}{X^2}$$

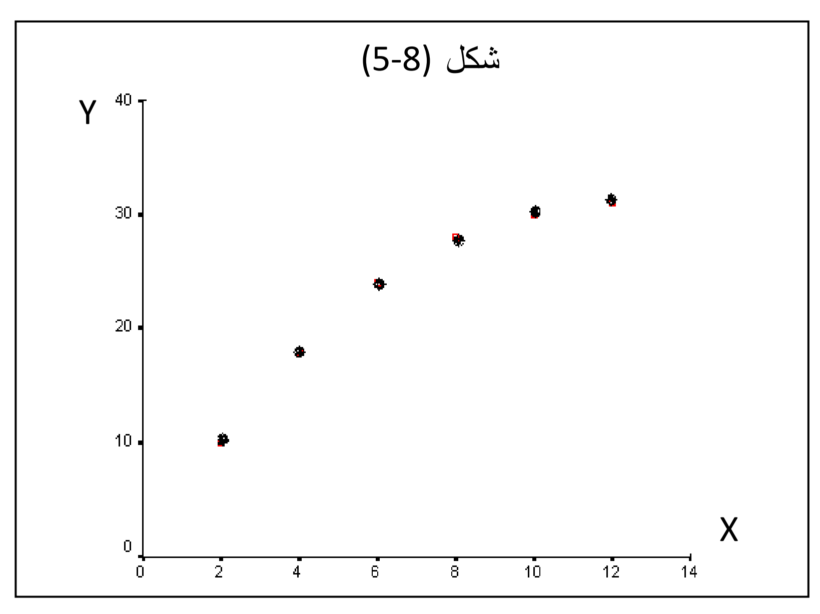
أي ان زيادة X بمقدار وحدة واحدة فان ذلك يقود الى انخفاض في المتغير Y بمقدار $\left(\frac{\beta_2}{X^2}\right)$ من الوحدات.

مثال(8-2): بيانات عن الكميات المنتجة Y الف طن وعدد العمال في أحد الصناعات X (الف عامل) لمدة (٦) سنوات متتالية:

جدول (8-4)

عدد العمال ألف عامل (X)	الكميات المنتجة ألف طن (Y)
2	10
4	18
6	24
8	28
10	30
12	31

الحل:



المصدر: برنامج SPSS لبيانات الجدول (8-4)

من خلال رسم الانتشار الشكل (8-5) يتضح ان الصيغة التي تمثل البيانات هي الصيغة اللوغارتمية المزدوجة، لذلك نستخدم نموذج الانحدار:

$$\ln Y = \beta_o + \beta_1 \ln X + u$$

وبحساب ln Y و ln X ثم تطبيق المربعات الصغرى يتم الحصول على معادلة التقدير والجدول(5-8) يتضمن الحسابات اللازمة لتقدير العلاقة.

$$l\, n\hat{Y} = 1.94 + 0.64 \ln X$$

$\hat{\beta}_1 = 0.64$ وهي تمثل مرونة الإنتاج بالنسبة للعمل.

اما اثر عدد العمال فى الكميات المنتجة (في المتوسط) فيتم حسابه على وفق الصيغة:

$$\frac{dY}{dX} = \beta_1 \frac{Y}{X}$$

حيث $\frac{Y}{X}$ تمثل $\frac{\bar{Y}}{\bar{X}}$ أو $\frac{\sum Y}{\sum X}$

جدول (٨–٥)

جدول الحسابات

$\ln Y = Y^*$	$\ln X = X^*$	$(\ln X)^2$	$(\ln X)(\ln Y)$
2.303	0.693	0.480	1.596
2.890	1.386	1.921	4.006
3.178	1.792	3.211	5.695
3.332	2.079	4.322	6.927
3.401	2.303	5.304	7.833
3.434	2.485	6.175	8.533

(٨–١–٤) <u>منحنى النمو اللوجستي: (Logistic Growth curve)</u>

ان النموذج المشهور الذي يستخدم لحساب معدل النمو للمتغيرات البايولوجية مثل نمو السكان أو انتشار التطور التقني فهو النموذج اللوجستي والذي يتمثل بالمعادلة .

$$Y_t = \frac{1}{1 + e^{-\beta_o - \beta_1 X}} \qquad \ldots \qquad (13-8)$$

Y_t: يمثل المتغير الذي يتم تبني تقنية جديدة في حال دراسة انتشار تقنية معينة مثل الحاسبات (أي ربات البيوت اللاتي يمتلكن حاسوب) ويكون X يمثل الزمن (t = 1,. . . ,t) أو عدد السكان لسنوات متتالية.

ويتضح من العلاقة (8–13) ان العلاقة غير خطية بدلالة المعلمات. ويتم تحويلها كالآتي:

$$Y_t + Y_t\, e^{-\beta_o - \beta_1 X} = 1$$

$$Y_t\, e^{-\beta_o - \beta_1 X} = 1 - Y_t$$

$$e^{-\beta_o-\beta_1 X}=\frac{1-Y_t}{Y_t}$$

$$-(\beta_o+\beta_1 X)=\ln(\frac{1-Y_t}{Y_t})$$

$$\beta_o+\beta_1 X \quad =\ln\frac{Y_t}{1-Y_t}$$

وبافتراض $\ln(\frac{Y_t}{1-Y_t})=Y_t^*$ وإضافة حد الاضطراب العشوائي فيصبح النموذج :

$$Y_t^*=\beta_o+\beta_1 X_t+u_t \qquad \cdots \qquad (14-8)$$

فالعلاقة (8–14) علاقة خطية بدلالة المتغيرات Y^* ، X. كما انها خطية بدلالة المعلمات β_0 و β_1 ويمكن تقديرها بموجب المربعات الصغرى الاعتيادية والصيغة (8–14) توضح ان معدل النمو يتزايد أولاً إلى نقطة محددة (وهي نقطة الانقلاب) والتي تظهر عندما $X=-\frac{\beta_0}{\beta_1}$ ، ثم يبدأ معدل النمو بالانخفاض . وبذلك فان المعلمة β_1 تسيطر على السرعة . وبالمقابل فان قيمة Y عند نقطة الانقلاب

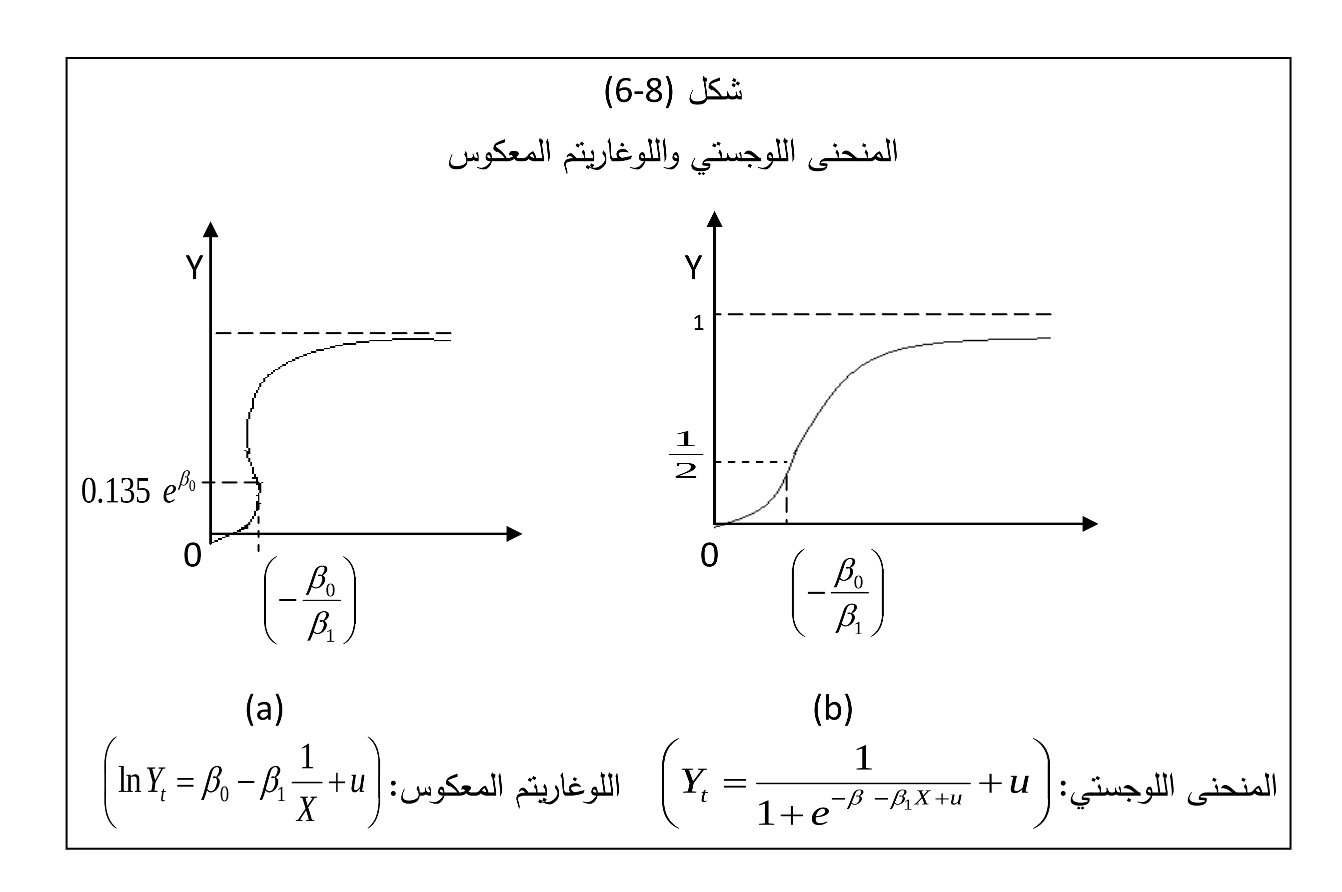
$$X=-\frac{\beta_0}{\beta_1}$$

$$\left(Y/X=-\frac{\beta_0}{\beta_1}\right)=\frac{1}{2}$$

وهناك تشابه كبير بين الصيغة اللوجستية وبين اللوغاريتم المعكوس وهي ان حد الإشباع (e^{β_0}).

ونقطة الانقلاب تظهر عندما $X=\frac{\beta_1}{2}$: $\left(Y/X=\frac{\beta_1}{2}\right)=0.135\ e^{\alpha}$ وتكون على بعد 13% من حد الإشباع.

والشكل (8–6) يوضح ذلك :



شكل (8-6)
المنحنى اللوجستي واللوغاريتم المعكوس

المنحنى اللوجستي: $\left(Y_t = \frac{1}{1+e^{-\beta-\beta_1 X+u}} + u\right)$ اللوغاريتم المعكوس: $\left(\ln Y_t = \beta_0 - \beta_1 \frac{1}{X} + u\right)$

وبشكل عام يمكن تحديد صيغ مختلفة التي يمكن ان تتصف بها العلاقة غير الخطية بين Y و X وذلك من خلال استخدام ما يسمى (محول بوكس – كوكس) Box- Cox transformation والتي تعتمد على الصيغة العامة:

$$Y^{\lambda_1} = \beta_o + \beta_1 X^{\lambda_2} + u \qquad \cdots \qquad (15-8)$$

بحيث :

$$Y^{\lambda_1} = \begin{cases} \frac{Y^{\lambda_1}-1}{\lambda_1} & \lambda_1 \neq 0 \quad \text{اذا} \\ \ln Y & \lambda_1 = 0 \quad \text{اذا} \end{cases}$$

وكذلك:

$$X^{\lambda_2} = \begin{cases} \frac{X^{\lambda_2}-1}{\lambda_2} & \lambda_2 \neq 0 \quad \text{اذا} \\ \ln X & \lambda_2 = 0 \quad \text{اذا} \end{cases}$$

مثال(8-3): اذا $\lambda_1 = \lambda_2 = 1$ فان الصيغة هي:

$$Y^{\lambda_1} = \frac{Y^1 - 1}{1} = (Y - 1) \qquad \Leftarrow \lambda_1 = 1 \neq 0$$

$$X^{\lambda_2} = \frac{X^1 - 1}{1} = (X - 1) \qquad \Leftarrow \lambda_2 = 1 \neq 0$$

وبذلك فان الصيغة بعد التعويض في العلاقة (8−15) نحصل على:

$$(Y - 1) = \beta_o + \beta_1 (X - 1) + u$$

$$Y = (1 + \beta_o - \beta_1) + \beta_1 X + u$$

$$Y = \beta^* + \beta_1 X + u \qquad \cdots \qquad (16\text{-}8)$$

أي

حيث : $\beta_o^* = 1 - \beta_o - \beta_1$

والعلاقة (8−16) هي علاقة خطية بدلالة المتغيرات X و Y وكذلك بدلالة بالمعلمات β^* و β_1.

مثال(8−٤): اذا كان $\lambda_1 = 0$, $\lambda_2 = 0$ فان العلاقة (8-15) تصبح الصيغة اللوغارتمية المزدوجة

$$\ln Y = \beta_o + \beta_1 \ln X + u$$

مثال(8-5): اذا كان $\lambda_1 = 0$, $\lambda_2 = -1$ فان تحويل بوكس – كوكس ينتج علاقة اللوغاريتم المعكوس log - reciprocal .

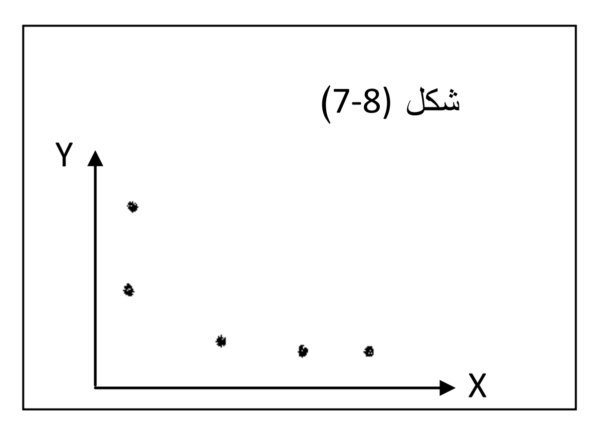
$$\ln Y = \beta_0^* + \beta_1^* \ln \frac{1}{X} + u$$

مثال(8-6): تقدير معدل التضخم (X) ومعدل البطالة (Y) لمجتمع خلال مدة (7) سبع سنوات.

جدول (8−6)

Y	X
20	4
16	6
14	8
13	10
12.5	12
12.25	14
12.125	16

رسم الانتشار يوضح ان الصيغة $Y = \beta_o + \beta_1 \frac{1}{X} + u$ هي الملائمة للبيانات



شكل (7-8)

جدول حسابات الانحدار للمثال (6-8)

جدول (8–7)

Y	X	$X^* = \frac{1}{X}$	$Y - \overline{Y}$	$X^* - \overline{X}^*$	$x^* y$	x^{*2}
20	4	0.25	5.732	0.127	0.7297	0.0162
16	6	0.167	1.732	0.044	0.07615	0.0019
14	8	0.125	-0.268	0.0023	-0.00062	0
13	10	0.1	٢٦٨-1.	-0.0227	0.02878	0.0005
12.5	12	0.083	.768١-	-0.0394	0.0696	0.00155
12.25	14	0.071	-2.018	-0.0513	0.1035	0.00236
12.125	16	0.063	-2.143	-0.0602	0.129	0.0036
Σ 99.875	70	0.859			1.13611	0.02638

$$\overline{X}^* = 0.1227 \;,\; \overline{X} = 10 \;,\; \overline{Y} = 14.268$$

$$\hat{\beta}_1 = \frac{1.13611}{0.02638} = 43$$

$$\hat{\beta}_0 = 14.268 - 43.(0.1227) = 8.99$$

وبذلك فان معادلة التقدير:

$$\hat{Y} = 8.99 + 43 \cdot \frac{1}{X}$$

فان زيادة معدل التضخم بنقطة واحدة سوف يصاحبها انخفاض في معدل البطالة بمقدار:

$-43\cdot\frac{1}{\bar{X}^2}=-43\cdot\frac{1}{(10)^2}=-0.43$ من النقطة في المتوسط.

وبذلك فان المرونة (مرونة البطالة للتضخم) $=\frac{-43}{10(14.268)}=-0.3$ وهي تشير الى ان الارتفاع في معدل التضخم بنسبة 10% يصاحبه انخفاض في معدل البطالة بنسبة 3% في المتوسط.

نماذج متعدد الحدود Polynomial

(٨-١-٥) <u>الصيغة التربيعية Quadratic</u>

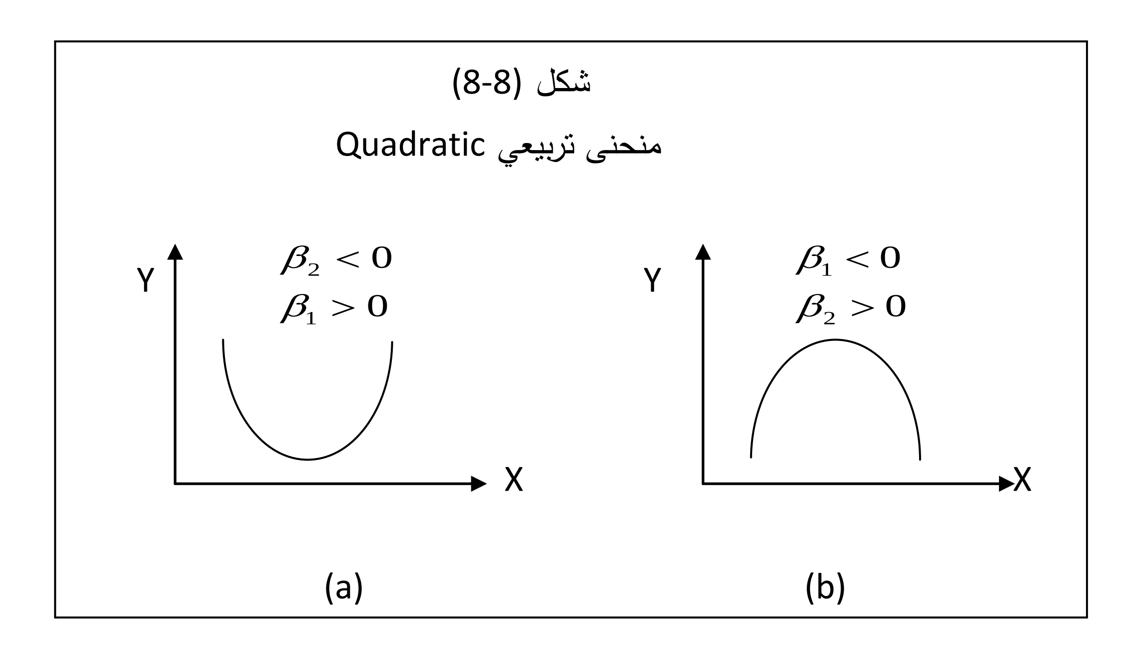
ان الصيغة العامة لمعادلة متعددة الحدود بدرجة k بدلالة متغير توضيحي واحد تكتب كالآتي:

$$Y=\beta_0+\beta_1X+\beta_2X^2+\ \cdots\ +\beta_kX^k+u \qquad \cdots \qquad (17\text{-}8)$$

وعندما تحدد درجة المعادلة هي الدرجة الثانية فان الصيغة تسمى (Quadratic)

$$Y=\beta_o+\beta_1X+\beta_2X^2+u \qquad \cdots \qquad (18\text{-}8)$$

ان شكل المنحني موضح في الشكل (8-8)



شكل (8-8)

منحنى تربيعي Quadratic

الصيغة (8-18) خطية بدلالة المعلمات ولكنها غير خطية بدلالة المتغيرات ، ولتحويلها الى خطية بدلالة المتغيرات ايضاً، يتم افتراض $X_1=X$ و $X_2=X^2$

فتصبح معادلة الانحدار:

$$Y=\beta_o+\beta_1X_1+\beta_2X_2+u$$

وهي معادلة انحدار خطي متعدد. ولأجل تقدير المعلمات بموجب β_0 و β_1 و β_2 يمكن تطبيق المربعات الصغرى الاعتيادية وعلى وفق القانون :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1} X'Y$$

حيث : $X'X = \begin{bmatrix} n & \sum X & \sum X^2 \\ \sum X & \sum X^2 & \sum X^3 \\ \sum X^2 & \sum X^3 & \sum X^4 \end{bmatrix}$ ، وان $X'Y = \begin{bmatrix} \sum Y \\ \sum XY \\ \sum X^2 Y \end{bmatrix}$

أو باستخدام المجاميع كانحرافات عن متوسطاتها:

$$\hat{\beta}_* = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (x'x)^{-1} x'y \quad \& \qquad \hat{\beta}_o = \bar{Y} - \hat{\beta}_1 \bar{X} - \hat{\beta}_2 \bar{X}^2$$

$$x'y = \begin{pmatrix} \sum xy \\ \sum x^2 y \end{pmatrix}, \quad x'x = \begin{pmatrix} \sum x^2 & \sum x^3 \\ \sum x^3 & \sum x^4 \end{pmatrix}$$

وهكذا فان الصيغة التكعيبية نحصل عليها عندما نعوض k=3 في المعادلة (8–١٧) فنحصل على:

$$Y = \beta_0 + \beta_1 X_1 \beta_2 X^2 + \beta_3 X^3 + u \qquad . \ . \ . \qquad (19-8)$$

وبالطريقة نفسها يتم تحويلها الى صيغة خطية.

بافتراض:

$$X = X_1$$

$$X^2 = X_2$$

$$X^3 = X_3$$

فتصبح العلاقة (8-19) نموذج انحدار خطي متعدد بدلالة المتغيراتX_1 ،X_2 ، X_3 .

وهكذا لكل نماذج الانحدار متعدد الحدود.

مثال(8-7): البيانات التالية خاصة بمعدل النمو الاقتصادي (X) والنصيب النسبي للطبقة الفقيرة من الدخل الكلي (Y) ، لعدد من الدول التي تختلف في مرحلة النمو الاقتصادي .

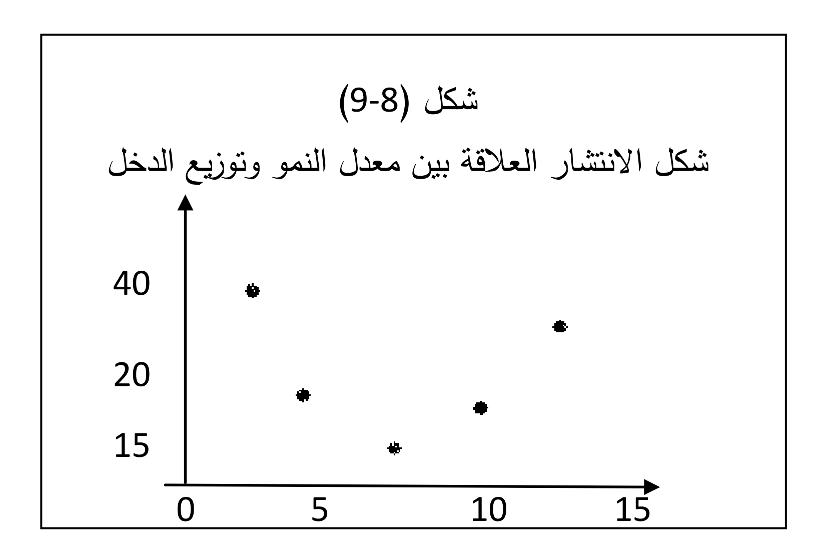
م/ تقدير العلاقة بين المتغيرين

جدول (8–8)

الدولة	١	٢	٣	٤	٥	٦	٧
معدل النمو %(X)	١	٣	٥	٧	٩	١١	١٣
النصيب النسبي (Y)	٣٥	٢٨	٢٢	١٧	٢١	٢٨	٣٨

من خلال رسم الانتشار الشكل (8-9) يتضح أن الصيغة التربيعية هي التي تمثل المشاهدات وبذلك تستخدم :

$$Y = \beta_o + \beta_1 X + \beta_2 X^2 + u$$



شكل (8-9)

شكل الانتشار العلاقة بين معدل النمو وتوزيع الدخل

جدول (٨–٩)

حسابات المثال (8-7)

	Y	X_1=X	X_2=X^2	y	x_1	x_2	yx_1	yx_2	x_1^2	x_2^2	x_1x_2
	٣٥	١	١	8	-6	-64	-48	-512	36	4096	384
	٢٨	٣	٩	1	-4	-56	-4	-56	16	3136	224
	٢٢	٥	٢٥	-5	-2	-40	10	200	4	1600	80
	١٧	٧	٤٩	-10	0	-16	0	١٦٠	0	256	0
	٢١	٩	٨١	-6	2	16	-12	-96	4	256	32
	٢٨	١١	١٢١	1	4	٥٦	4	56	16	3136	224
	٣٨	١٣	١٦٩	11	6	104	66	١١٤٤	36	10816	624
Σ	٨٩	٤٩	٤٥٥				١٦	٨٩٦	١١٢	٢٣٢٩٦	١٥٦٨

$$\bar{X}_1=\frac{49}{7}=7 \quad , \qquad \bar{X}_2=\frac{455}{7}=65 \qquad , \qquad \bar{Y}=\frac{189}{7}=27$$

اذن المعادلات الطبيعية للقيم كانحرافات عن متوسطاتها :

$$16=112\hat{\beta}_1+1568\hat{\beta}_2$$
$$896=1568\hat{\beta}_1+23296\hat{\beta}_2$$

وبحل المعادلتين ينتج : $\Rightarrow \quad \hat{\beta}_1=-6.857 \quad , \quad \hat{\beta}_2=0.5$

تعوض القيم في المعادلة: $\hat{\beta}_o=\bar{Y}-\hat{\beta}_1\bar{X}_1-\hat{\beta}_2\bar{X}_2$ للحصول على المقطع الصادي $\hat{\beta}_o=42.5$

وبذلك فان معادلة الانحدار: $\hat{Y}=42.5-6.857X+0.5X^2$

ومن معادلة التقدير نجد ان النصيب النسبي للطبقة الفقيرة من الدخل قبل بدء عملية النمو (عندما يكون معدل النمو = صفر) يسأوي ٤٢.٥ % في المتوسط . وبذلك فان النصيب النسبي للطبقة الغنية هو % ٥٧.٥ في المتوسط .

وللحصول على الحد الأدنى للنصيب النسبي للطبقة الفقيرة من الدخل ، نفاضل معادلة الانحدار ونجعلها = صفر

$$\frac{d\hat{Y}}{dX}=-6.857X$$
$$\Rightarrow X=6.857$$

وبتعويض القيمة في معادلة الانحدار:

$$\left(\hat{Y}/X=6.857\right)=42.5-6.857(6.857)+0.5(6.857)^2$$
$$=18.98$$

أي أن الحد الأدنى لنصيب الطبقة الفقيرة هو ١٩% تقريبا وللتأكد من أن هذه القيمة تمثل نهاية صغرى نتحقق من ذلك بالمشتقة الثانية :

$$\frac{d^2\hat{Y}}{dX^2}=-6.857<0$$

اذن النقطة هي نهاية صغرى.

اختبار أهمية المتغير X في نماذج متعددة الحدود.

ان الامر الذي سيتم توضيحه في هذه الفقرة بشأن اختبار الأهمية للمتغيرات التوضيحية في معادلة الانحدار. ان اختبار معنوية المعلمات بشكل عام في نماذج الانحدار غير الخطي لا يختلف كثيرا عنها في معادلة الانحدار الخطي، فالاختلاف يكمن في تفسير المعلمة وكما أوضحنا في الفقرات السابقة. فان معلمة الانحدار في نموذج اللوغاريتم المزدوج تمثل مرونة Y بالنسبة الى X ، وفي النموذج شبه اللوغاريتم لـ (Log – Lin) Y تمثل معدل النمو اللحظي ومنها يحسب معدل النمو المركب للمتغير Y اذا كان X يمثل الزمن.

وهكذا في النماذج غير الخطية الأخرى والجدول التالي يبين اثر المتغير X فى Y والمتمثل $\left(\frac{dy}{dx}\right)$ في انماط المعادلات غير الخطية المختلفة.

جدول (8–10)

العلاقة	الصيغة	الأثر: القيمة الحدية $\frac{dY}{dX}$
الخطية	$Y = \beta_o + \beta\ X$	β_1
اللوغاريتم المزدوج Log-log	$\ln Y = \beta_o + \beta_1 \ln X$	$\beta_1 \cdot \frac{Y}{X}$
شبه اللوغاريتمLog-lin	$\ln Y = \beta_o + \beta_1 X$	$\beta_1 Y$
Lin-log	$Y = \beta_0 + \beta_1 \ln X$	$\beta_1 \cdot \frac{1}{X}$
العكسية	$Y = \beta_0 + \beta_1 \frac{1}{X}$	$-\beta_1 \cdot \frac{1}{X^2}$
لوغاريتم العكسي	$\log Y = \beta_0 + \beta_1 \frac{1}{X}$	$-\beta_1 \cdot \frac{Y}{X^2}$
التربيعية	$Y = \beta_o + \beta_1 X + \beta_2 X^2$	$\beta_1 + 2\beta_2 X$
التكعيبية	$Y = \beta_o + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$	$\beta_1 + 2\beta_2 X + 3\beta_3 X^2$

أما اختبار معنوية المعلمات المقدرة $(\hat{\beta}_1)$ فيعني معنوية المتغير المرافق (X^*) في العلاقة المعنية نسبة الى (Y^*) ففي الصيغة اللوغاريتمية المزدوجة Log .Log فان $(\hat{\beta}_1)$ تمثل مرونة Y بالنسبة

لـ $(Y^{*}=\log Y$ و X$X^{*}=\log X)$، وهكذا معنوية $(\hat{\beta}_1)$ في العلاقة العكسية تعني معنوية المتغير $\left(\frac{1}{X}\right)$ بالنسبة لـ Y، وفي علاقة اللوغاريتم العكسي تدل على معنوية المتغير $\left(\frac{1}{X}\right)$ على log Y.

أما بالنسبة لنماذج متعدد الحدود فان الحالة تختلف قليلاً. ولتوضيح ذلك نفترض نموذج الانحدار متعدد الحدود:

$$Y=\beta_0+\beta_1X+\beta_2X^2+...+\beta_kX^k+u$$

فعند اختبار معنوية انحدار جزئي معين (مثلا أختبار معنوية β_1) فتكون فرضية العدم:

$$H_0:\beta_1=0$$

ومع افتراض ان جميع المعلمات بدرجات أعلى تسأوي صفراً أي: $\beta_2=\beta_3=...=\beta_k=0$

اما في حالة الانحدار الخطي المتعدد :

$$Y=\beta_0+\beta_1X+\beta_2X_2+\beta_3X_3+...+\beta_kX_k+u$$

فان اختبار معنوية (β_1) لايعني تجاهل أو اهمال $\beta_k,...,\beta_3,\beta_2$ فالغاية من الاختبار $(\beta_1=0)$ في الانحدار الخطي المتعدد هل ان اضافة X_1 الى المعادلة سيساعد معنويا في التنبؤ لمعدل Y بوجود غيره من المتغيرات X_2، X_3، . . . ، X_k. في حين تكون الغاية من اختبار $(\beta_1=0)$في الانحدار متعدد الحدود هو لتحديد درجة المعادلة.

والمثال التالي يوضح الاختلافات التي نود توضيحها من خلال مقارنتها مع نماذج الانحدار المتعدد الخطي.

مثال (٨-٨): البيانات في الجدول : (٨-١١) تمثل مقدار السكر المتحول (Y) في عملية كيميأوية معينة عند درجات حرارة مختلفة (X_i).

جدول (8-11)

Y	6	6.2	6.8	7.5	8.6	9.7	11.6	14.1
x	6	7	8	9	10	11	12	13

ولتوفيق معادلة من الدرجة الثالثة والتي صيغتها الرياضية: $Y=\beta_0+\beta_1X+\beta_2X^2+\beta_3X^3+u$

فيتم تنظيم البيانات كما يلي:

جدول (8-١٢)

المشاهدات	Y	X	$X_1^* = X$	$X_2^* = X^2$	$X_3^* = X^3$
١	6	6	6	36	216
2	6.2	7	7	49	343
3	6.8	8	8	64	512
4	7.5	9	9	81	729
5	8.6	10	10	100	1000
6	9.7	11	11	121	1331
7	11.6	12	12	144	1728
8	14.1	13	13	169	2197
Σ	70.5	76			

ومن حسابات الجدول (8-١٠) فان المعادلات الطبيعية للنموذج متعدد الحدود من الدرجة الثالثة :

$$\begin{bmatrix} n & \Sigma X_1^* & \Sigma X_2^* & \Sigma X_3^* \\ & \Sigma X_2^{*2} & \Sigma X_1^* X_2^* & \Sigma X_1^* X_3^* \\ & & \Sigma X_2^{*2} & \Sigma X_2^* X_3^* \\ & & & \Sigma X_3^{*2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1^* Y \\ \Sigma X_2^* Y \\ \Sigma X_3^* Y \end{bmatrix}$$

$$\begin{bmatrix} 8 & 76 & 764 & 8056 \\ & 764 & 8056 & 88292 \\ & & 88292 & 9979576 \\ & & & 11542244 \end{bmatrix} \begin{bmatrix} \beta_o \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 70.5 \\ 716.5 \\ 7649.5 \\ 84904.9 \end{bmatrix}$$

وبذلك فان المعلمات المقدرة:

$$\begin{bmatrix} \hat{\beta}_o \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0.0154 \\ 0.2693 \\ 1.9180 \\ 0.8043 \end{bmatrix}$$

أي ان معادلة الانحدار التقديرية: $\hat{Y} = 0.0154 + 0.2693X + 1.9180X^2 + 0.8043X^3$

ولاختبار هل هناك انحدار معنوي عام يجب حساب جدول تحليل التباين (ANOVA)

باختبار فرضية العدم:

$$H_o = \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_1 = \beta_1 \neq \beta_2 \neq \beta_3 \neq 0$$

ينظم جدول تحليل التباين في جدول (8–13)

جدول (8–13)

s.o.v	d.f	SS	MS	F
$Regression(X_1^* X_2^* X_3^*)$	3	57.0129	19.0043	1367.2158
$Error(X_1^* X_2^* X_3^*)$	4	0.0557	0.0139	
$Total$	7	57.0068		

وباختبار مستوى دلالة 1% فان قيمة F الجدولية $F_{(3,4,0.99)} = 16.69$ وبمقارنة F المحسوبة من جدول تحليل التباين جدول (8–11) وهي (١٣٦٧.٢) مع F الجدولية (١٦.٦٩) فيكون القرار برفض H_0 أي ان هناك انحدار عام باستخدام المتغيرات X_3^*, X_2^*, X_1^* .

– ولاختبار هل ان أضافة X_2^* سيساعد في تنبؤ Y فان فرضية العدم: $H_o : \beta_2 = 0$

على افتراض ان β_3 تسأوي صفراً. (أي ضمنياً هذه الفرضية تفترض ان Y تتحدد بـ X_2^*, X_1^* فقط) فالمختبر الاحصائي المستخدم في هذه الحالة:

$$F^* = \frac{ESS(X_2^* / X_1^*)/1}{RSS(X_1^*, X_2^*)/(n-3)}$$

وبذلك: $ESS(X_2^* / X_1^*) = ESS(X_1^*, X_2^*) - ESS(X_1^*)$

حيث كان $ESS(X_1^*)$ يمكن ايجاده من علاقة الانحدار البسيط : $Y = \beta_o + \beta_1 X_1^* + u$

$$ESS(X_1^*) = \hat{\beta}_1 S_{X_1^* Y}$$

في حين $\hat{\beta}_1$ يمكن حسابها :

$$\hat{\beta}_1 = \frac{S_{X_1^* Y}}{S_{X_1^* X_1^*}}$$

$$S_{X_1^* Y} = \sum X_1^* Y - \frac{\sum X_1^* \sum Y}{n} = 716.5 - \frac{(76)(70.5)}{8} = 46.75$$

$$S_{X_1^*X_1^*} = \sum X_1^{*2} - \frac{(\sum X_1^*)^2}{n} = 764 - \frac{(76)^2}{8} = 42$$

$$\hat{\beta}_1 = \frac{46.75}{42} = 1.1130$$

$$ESS(X_1) = \hat{\beta}_1\ S_{X_1^*Y} = 1.1130(46.75) = 52.0372$$

بينما يمكن حساب $ESS(X_1^* X_2^*)$ من الانحدار: $\quad Y = \beta_o + \beta_1 X_1^* + \beta_2 X_2^* + u$

حيث المعادلات الطبيعية:

$$\begin{pmatrix} n & \Sigma X_1^* & \Sigma X_2^* \\ & \Sigma X_1^{*2} & \Sigma X_1^* X_2^* \\ & & \Sigma X_2^{*2} \end{pmatrix} \begin{pmatrix} \beta_o \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma X_1^* Y \\ \Sigma X_2^* Y \end{pmatrix}$$

وبحل المعادلات نحصل على المقدرات:

$$\begin{pmatrix} \hat{\beta}_o \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1} X'Y = \begin{pmatrix} 12.657 \\ -2.110 \\ 0.169 \end{pmatrix}$$

وبذلك فان

$$ESS(X_1^* X_2^*) = \hat{\beta}' X'Y - \frac{(\Sigma Y)^2}{n} = 56.8702$$

وعليه فان $ESS(X_2^* / X_1^*)$ يمكن حسابه كالآتي:

$$ESS\left(X_2^* / X_1^*\right) = 56.8702 - 52.0372 = 4.833$$

في حين $RSS\left(X_1^* X_2^*\right)$ يمكن حسابه:

$$RSS\left(X_1^* X_2^*\right) = TSS - ESS\left(X_1^* X_2^*\right) = 0.1967$$

وبذلك فان قيمة F الحسابية لاختبار أهمية اضافة X_2^* هي:

$$F^* = \frac{4.833}{0.196/5} = \frac{4.833}{0.03934} = 122.85$$

وبمقارنتها بالقيمة الجدولية: $\quad F_{C(1,5,0.99)} = 16.26$

فيكون القرار: رفض فرضية العدم: بمعنى ان اضافة الحد X^2 الى معادلة الخط المستقيم يساعد في التنبؤ لقيم Y

– اما اذا كان الهدف هو اختبار الفرضية: $H_0:\ Y=\beta_2 X^2$

فتصاغ الفرضية كالآتي: $H_0:\beta_0=\beta_1=\beta_3=0$

وبذلك فان خطوات الاختبار تكون .

١– نحسب مجموع المربعات المشروحة الاجمالية غير المصححة :

$$ESS\left(X_o X_1{}^* X_2{}^* X_3{}^*\right)=\hat{\beta}'X'Y$$

$$=678.294$$

٢– نحسب مربعات الانحدار غير المصححة لـ $X_2{}^*$:

$$ESS\left(X_2{}^*\right)=\hat{\beta}_2\cdot\Sigma X_2 Y$$

حيث $\hat{\beta}_2$ تحسب من الانحدار البسيط $Y=\beta_o+\beta_2 X_2$ على وفق الأتي:

$$\hat{\beta}_2=\frac{7649.5}{88292}=0.0866$$

$$ESS\left(X_2{}^*\right)=(0.0866)(7649.5)=662.742$$

وبذلك فان مجموع المربعات المشروحة لفرضية العدم هي:

$$ESS\left(X_0 X_1{}^* X_3{}^*/X_2{}^*\right)=ESS\left(X_0\ \ X_1{}^* X_2{}^* X_3{}^*\right)-ESS\left(X_2{}^*\right)=15.552$$

وبذلك فان جدول تحليل التباين لاختبار الفرضية : $H_0:Y=\beta_2 X^2$

كما موضح في جدول (8–12)

جدول (8–12)

جدول تحليل التباين لاختبار : $H_0:Y=\beta_2 X^2$

s.o.v	d.f	SS	MS	F
$\text{Re}\,g\left(X_0\ \ X_1{}^* X_2{}^* X_3{}^*\right)$	4	678.294		
$\text{Re}\,g\left(X_2{}^*\right)$	1	662.742		
$\text{Re}\,g\left(X_0 X_1{}^* X_3{}^*/X_2{}^*\right)$	3	15.552	5.184	370.28
Error	4	0.056	0.014	
Total	8	678.35		

وبذلك فان القرار هو رفض فرضية العدم. بعبارة اخرى ان $(\beta_0\neq\beta_1\neq\beta_3\neq 0)$

أي أن الصيغة $Y\neq\beta_2 X^2$.

أسئلة الفصل الثامن

س1: 1. وضح الدلالة الإحصائية للعبارة التالية:

الخطية في نموذج الانحدار.

2. بين أي من النماذج التالية هي علاقة انحدار خطية.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$$

$$\ln Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

$$\frac{1}{Y} = \beta_0 + \beta_1 X + \beta_2 \sqrt{X} + u$$

$$Y = \beta_0 + \sqrt{\beta_1} X + \beta_2 \ln X + u$$

$$Y = \beta_0 + \beta_1 \beta_2 X + u$$

س2: 1. حدد الصيغة العامة لنموذج النمو الثابت.

2. احسب معدل النمو المركب لبيانات الناتج المحلي الاجمالي العراقي (GDP) بملايين الدنانير عبر السنوات (2004-2010).

السنوات	2004	2005	2006	2007	2008	2009	2010
GDP	38058543.0	53386428.6	80459422.4	93981672.4	157026061.6	130642187.0	158521511.5

س3: اذا علمت ان معادلة تقدير الصادرات (Y) كدالة نصف لوغاريتمية بدلالة الناتج المحلي الاجمالي (X) اعطت الصيغة التالية.

$$\hat{Y} = -12.5 + 0.85 \ln X$$

أ) فسر معلمة الناتج المحلي الاجمالي.

ب) ما هو أثر الناتج المحلي الاجمالي فى قيمة الصادرات؟

س4: وضح أهم اشكال العلاقة العكسية بالرسم، ثم اذكر أهم التطبيقات التي يمكن اعتمادها لكل شكل، موضحاً القيود الممكنة لمعلمات كل شكل منها.

س5: في علاقة انحدار الكميات المنتجة من الجبن في مصنع معين (Y) بدلالة عدد العمال (X) تم الحصول على الاتي:

$$\ln \hat{Y} = 3.5 + 0.8 \ln X$$

أ) فسر المدلول للمعلمات المقدرة.

ب) ما أثر عدد العمال فى الكميات المنتجة من الالبان؟

س6: على وفق تحويل بوكس–كوكس بين الصيغة الملائمة للحالات التالية.

$$\lambda_2 = \frac{1}{2} \quad , \quad \lambda_1 = 0 \qquad (1)$$

$$\lambda_2 = -1 \quad , \quad \lambda_1 = 1 \qquad (2)$$

$$\lambda_2 = 0 \quad , \quad \lambda_1 = -1 \qquad (3)$$

$$\lambda_2 = 0 \quad , \quad \lambda_1 = 1 \qquad (4)$$

س7: اذا علمت ان نصيب النسبي للطبقة الفقيرة من الدخل الكلي (Y) ومعدل النمو الاقتصادي (X) لعينة من الاقطار المختلفة قد اعطت الصيغة التقديرية.

$$\hat{Y} = 39.7 - 5.32X + 0.3X^2$$

أ) احسب النصيب النسبي للطبقة في المتوسط.

ب) احسب الحد الادنى للنصيب النسبي للطبقة الفقيرة من الدخل.

س8: وضح أهم الفروق الرئيسة لاختبار الفرضيات في النموذج الخطي لمتعدد المتغيرات مع الانحدار غير الخطي البسيط.

الفصل التاسع

اختيار أحسن المعادلات Selecting the best regression equation

سيتم في هذا الفصل التعريف بأهم الطرائق التي تستخدم من اجل اختيار أحسن معادلة خطية مؤلفة من متغير معتمد واحدY، وعددمن المتغيرات المستقلة (التوضيحية) $(X_1, X_2, \ldots ..X_k)$.

من المشكلات التي تواجه الباحث في مجال تطبيق نظام معادلات الانحدار مشكلة وجود عدد كبير من المتغيرات المفسرة في حين يرغب الباحث في اختيار أفضل مجموعة من هذه المتغيرات لتكوين نموذج أمثل يعبر عن هذا النظام نظراً لان إدخال إعدادا كبيرة من المتغيرات المستقلة في المعادلة يكلف ثمناً كبيراً، مما يكلفه من جهد، ووقت ومال .

ويعرف النموذج الأمثل بأنة النموذج الذي ينتج عن اختيار متغيراته المفسرة على وفق معايير عدة

1-أدنى قيمة لمحدد مصفوفة متوسطات أخطاءالتنبؤ أو أقل قيمة لمتوسط مربعات الخطأ :

$$\frac{RSS}{n-k-1}$$

2- معيارمعامل التحديد . R^2 أومعيار $\bar{R}^2$ (معامل التحديد المعدل) أعلى قيمة (اكبر من 60%) اذ يتم اختيار النموذج الذي يمتلك أعلى قيمة لـR^2 . أو $\bar{R}^2$

3- معيار C_P (احصاءة مالوMallows static)

$$C_p = \frac{RSS(X_1, \cdots, X_p)}{\dfrac{RSS(X_1, \cdots, X_k)}{(n-k-1)}} - (n-2p)$$

حيث:k عددالمتغيرات المستقلة جميعها .

Pعدد معلمات النموذج المقدر بضمنها المقطع الصادي .

ويتم اختيار النموذج الذي يمتلك أقل قيمة لـ C_P.

وهناك العديد من طرائق الاختيار وسيتم التركيز على ثلاث طرائق منها وهي :

طريقة الحذف العكسي أو الخلفي The Backward elimination procedure

طريقة الاختيار الأمامي أو المباشر The Forward Selection procedure

طريقة الاختيار التدريجي The stepwise selection procedure

ولابد من التأكيد على إن النتائج قد تختلف باستخدام طريقة دون أخرى .

(1−9) طريقة الحذف العكسي أو الخلفي : The Backward elimination procedure

تتلخص الطريقة بالخطوات التالية :

– نبدأ باعتماد جميع المتغيرات المستقلة في المعادلة ثم تحذف المتغيرات من المعادلة واحداً بعد الأخر اعتماداً على قيمة (F) الجدولية والتي تسمى (Fout) وسيتم تفصيل الخطوات على وفق الأتي:

الخطوة الأولى:

يتم العمل على تضمين جميع المتغيرات المستقلة في معادلة الانحدار وتحسب قيم (F) الجزئية لكل متغير (F الإضافية) على وفق الصيغة التالية :

$$F_i = \frac{ESS(X_i / all\ other\ \exp lanatory\ \ \text{var}\, iables)}{RSS(X_1, \cdots, X_k)/(n-k-1)} \qquad \forall \qquad i = 1, \cdots, k$$

ويختار المتغير الذي له (أقل قيمة) F جزئية (F_i) وبعد مقارنتها مع (Fout).

فإذا اثبت إن $F_i < Fout$ يتم حذف المتغير المعني من المعادلة . ويتم الانتقال إلى الخطوة الثانية .

حيث إن (Fout) قيمة جدوليه بدرجات حرية البسط واحد والمقام $(n-k-1)$ وبمستوى دلالة **%** α.

الخطوة الثانية:

المتغيرات التوضيحية عدا الذي تم حذفه في الخطوة الأولى.

يتم احتساب (F) الجزئية لكل متغير من المتغيرات الباقية من الخطوة الأولى ، ويتم اختيار أصغرها، وتقارن مع (Fout) بدرجات حرية البسط (1) والمقام (n-k-2)

فإذا كانت (F الأصغر) الجزئية المحسوبة (اصغر) من (Fout) يحذف المتغير المعني ويتم الانتقال إلى الخطوة الثالثة وهكذا تستمر الخطوات إلى إن يتم الحصول على إن **(اصغر) قيمة (F) جزئية تكون (اكبر) من (Fout) فيتوقف الحل.**

ولا بد من التأكيد انه من الضروري احتساب مقاييس المفاضلة R_p^2 ، $C_p = \frac{RSS}{n-p-2}$ لكل خطوة من خطوات الحل.

مثال(9−1): عينة بحجم (25) مشاهدة للمتغير المعتمد y مع المتغيرات المستقلة X_1 ، X_2 ، X_3 ، X_4
وباستخدام طريقة المربعات الصغرى أعطت المعلومات التالية : $TSS = 63.77$

المتغيرات	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4
ESS	14.35	1.19	18.33	9.31	20.15	26.61	22.07	19.34	10.89	18.69

$X_1X_2X_3$	$X_1X_2X_4$	$X_1X_3X_4$	$X_2X_3X_4$	$X_1X_2X_3X_4$
30.9	26.35	27.33	19.81	32.1

م/ إيجاد أفضل معادلة انحدار باستخدام طريقة الحذف العكسي :

الخطوة الأولى :

$$F_1^* = \frac{ESS(X_1/X_2X_3X_4)}{RSS(X_1X_2X_3X_4)/(n-5)}$$

$$ESS(X_1/X_2X_3X_4) = ESS(X_1X_2X_3X_4) - Ess(X_2X_3X_4) = 32.1 - 19.81 = 12.29$$

$$RSS(X_1X_2X_3X_4) = Tss - Ess(X_1X_2X_3X_4) = 31.67$$

$$F_1^* = \frac{12.29}{1.5835} = 7.76 \quad , \quad F_{c(1,20,0.95)} = 4.35$$

وكذلك نحسب F_2^* و F_3^* و F_4^* كما في الجدول :

S.O.V	d.f	SS	MS	Fi	Fout(1,20,0.95)
$E(X_1X_2X_3X_4)$	4	32.1			4.35
$E(X_1/X_2X_3X_4)$	1	32.1−19.81=12.29	12.29	7.78	
$E(X_2/X_1X_3X_4)$	1	32.1−27.33=4.77	4.77	3.02	
$E(X_3/X_1X_2X_4)$	1	32.1−26.35=5.75	5.75	3.64	
$E(X_4/X_1X_2X_3)$	1	32.1−30.9=1.2	1.2	0.76	الأصغر←
$R(X_1X_2X_3X_4)$	20	63.77−32.1=31.67	1.58		
Total	24				

الاصغر $F_4^* = 0.76 < \underset{4.35}{Fout}$

← X_4 يحذف من المعادلة . ← $Y = f(X_1X_2X_3)$

الخطوة الثانية :

S.O.V	d.f	SS	MS	F_i^*	Fout(1,21)
$E(X_1X_2X_3)$	3	30.9			4.32
$E(X_1/X_2X_3)$	1	30.9−19.34=11.56	11.56	7.36	
$E(X_2/X_1X_3)$	1	30.9−26.61=4.29	4.29	2.73	الأصغر ←
$E(X_3/X_1X_2)$	1	30.9−20.15=10.75	10.75	6.85	
$R(X_1X_2X_3)$	21	63.77−30.9=32.87	1.57		
Total	24				

الأصغر F_2^* أقل من Fout ← يحذف المتغير X_2 من المعادلة .

$$Y = f(X_1, X_3)$$

الخطوة الثالثة :

S.O.V	d.f	SS	MS	F_i^*	Fout(1,22,0.95)
$E(X_1X_3)$	2	26.61			4.3
$E(X_1/X_3)$	1	26.61−18.33=8.28	8.28	4.9	الأصغر ←
$E(X_3/X_1)$	1	26.61−14.35=12.26	12.26	7.25	
$R(X_1X_3)$	22	63.77−26.61=37.16	1.69		
Total	24				

المتغير X_1 له أقل قيمة لـF وهي اكبر من Fout

إذن أفضل معادلة انحدار هي :

$$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + u$$

$$F^* = \frac{ESS(X_1X_3)/2}{RSS(X_1X_3)/22} = \frac{13.305}{1.69} = 7.873$$

$$R^* = \frac{26.61}{63.77} = 0.42 \qquad \Rightarrow \qquad F^* \rangle\ F_{c(1,20,0.95)}$$

$$F^{*} > F_{(1,22,0.95)}$$

مثال(2−9):عينة بحجم (13) مشاهدة للمتغير المعتمد Y مع المتغيرات المستقلة X_1 ، X_2 ،X_3 ،X_4،
مع توافر المعلومات التالية:

المتغيرات	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3
ESS	61.476	404.327	558.054	151.844	416.43	559.018	210.828	625.515

↓

X_2X_4	X_3X_4	$X_1X_2X_3$	$X_1X_2X_4$	$X_1X_3X_4$	$X_2X_3X_4$	$X_1X_2X_3X_4$
548.177	871.836	636.168	560.708	872.029	900.054	910.841

$$TSS = 1388$$

الحل: بطريقة الحذف العكسي:

الخطوة الأولى :

S.O.V	d.f	SS	MS	F_i^*	Fout(1,8,0.95)
$E(X_1X_2X_3X_4)$	4	910.841			5.32
$E(X_1/X_2X_3X_4)$	1	910.841− 900.054=10.787	10.787	0.1808	الأصغر ←
$E(X_2/X_1X_3X_4)$	1	910.841− 872.029=38.812	38.812	0.651	
$E(X_3/X_1X_2X_4)$	1	910.841−560.708=350.133	350.133	5.870	
$E(X_4/X_1X_2X_3)$	1	910.841− 636.168=274.673	274.673	4.605	
$R(X_1X_2X_3X_4)$	8	1388 − 910.841= 477.159	59.645		
Total	12	1388			

هي الأصغر وهي أقل من Fout ← X_1يحذف من المعادلة

$$Y = f_1(X_2X_3X_4)$$

الخطوة الثانية:

S.O.V	d.f	SS	MS	F_i^*	Fout(1,9,0.95)
$E(X_2X_3X_4)$	3	900.054			5.12
$E(X_2/X_3X_4)$	1	900.054 − 871.836 = 28.218	28.218	0.52	الأصغر ←
$E(X_3/X_2X_4)$	1	900.054 − 548.177 = 351.877	351.877	6.49	
$E(X_4/X_2X_3)$	1	900.054 − 625.515 = 274.539	274.539	5.06	
$R(X_2X_3X_4)$	9	1388 − 900.054 = 487.946	54.216		
Total	12	1388			

F_2^* هي الأصغر وهي اقل من Fout← X_2 يحذف من المعادلة .

$$Y = f_2(X_3X_4)$$

الخطوة الثالثة:

S.O.V	d.f	SS	F_i^*	Fout
$E(X_3X_4)$	2	871.836		4.96
$E(X_3/X_4)$	1	871.836 − 151.844 = 719.992	13.95	
$E(X_4/X_3)$	1	871.836 − 558.054 = 313.782	6.08	الأصغر ←
$R(X_3X_4)$	10	1388 − 871.836 = 516.164		

F_4^* هو الأصغر وله قيمة اكبر من Fout

وبذلك فان أفضل معادلة انحدار هي : $Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + u$

$$F^* = \frac{ESS(X_3X_4)/2}{RSS(X_3X_4)/10} = \frac{435.918}{51.6164} = 8.445$$

$$R^* = \frac{871.836}{1388} = 0.63 \quad \Rightarrow \quad F^* \rangle F_{c(1,20,0.95)}$$

مثال(9−3): استخدم اسلوب الحذف العكسي مع توافر المعلومات التالية:

TSS =146 ، n=20

ESS	53	63	13	66	70	79	81
Var	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$

الخطوة الأولى: نحسب ESS

S.O.V	d.f	SS	MS	F_i^*	Fout(1,16,0.95)
$E(X_1X_2X_3)$	3	81			4.49
$E(X_1/X_2X_3)$	1	81−79 = 2	2	0.492	الأصغر ←
$E(X_2/X_1X_3)$	1	81−70 = 11	11	2.708	
$E(X_3/X_1X_2)$	1	81− 66 = 15	15	3.692	
$R(X_1X_2X_3)$	16	65	4.0625		

الأصغر $Fout \rangle F_{X_1}^*$ ← X_1 يحذف من المعادلة .

$$Y = f(X_2,X_3)$$

الخطوة الثانية:

S.O.V	d.f	SS	MS	F_i^*	Fout(1,17,0.95)
$E(X_2X_3)$	2	79			4.45
$E(X_2/X_3)$	1	79 −13 = 66	66	16.9	
$E(X_3/X_2)$	1	79 – 63 = 16	16	4.10	الأصغر ←
$R(X_2X_3))$	17	67	3.9		

الأصغر $Fout < F_3^*$

يحذف X_3 من المعادلة

إذن $Y = f(X_2)$

الخطوة الثالثة :

$$F_2^* = \frac{ESS(X_2)}{RSS(X_2)/n-2} = \frac{63}{83/18} = \frac{63}{4.6} = 14$$

Fout (1 , 18 , 0.95) = 4.41

$$F_2^* > Fout$$

أذن أفضل معادلة انحدار هي :

$$Y = \beta_0 + \beta_2 X_2 + u$$

مثال(9−4): الجدول يعرض مجموعة مربعات أخطاء العلاقة باستخدام خيارات مختلفة لعينة (n=13)

$TSS = 11.058$

المتغيرات	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	$X_1X_2X_3$	$X_1X_2X_4$	$X_2X_3X_4$	$X_1X_3X_4$	$X_1X_2X_3X_4$
RSS	0.607	10.795	10.663	1.522	0.499	0.600	0.580	10.168	1.218	0.498	0.450	1.041	0.581	0.441

م/استعمل طريقة الحذف العكسي لتقدير أفضل معادلة انحدار

الخطوةالأولى :

s.o.v	d.f	SS	MS	F^*
$E(X_1/X_2X_3X_4)$	1	0.6	0.6	10.884
$E(X_2/X_1X_3X_4)$	1	0.14	0.14	2.539
$E(X_3/X_1X_2X_4)$	1	0.009	0.009	0.1633 ← الأصغر/
$E(X_4/X_1X_2X_3)$	1	0.057	0.057	1.034
$RSS(X_1X_2X_3X_4)$	8	0.441	0.055125	

$$ESS(X_1/X_2X_3X_4) = ESS(X_1X_2X_3X_4) - ESS(X_2X_3X_4)$$

$$= SST - RSS(X_1X_2X_3X_4) - SST + RSS(X_2X_3X_4)$$

$$= RSS(X_2X_3X_4) - RSS(X_1X_2X_3X_4)$$

$Fout\ : F_c(1,8,0.95) = 5.32$ ← X_3 يحذف

$$Y = f(X_1X_2X_4)$$

الخطوة الثانية:

s.o.v	d.f	SS	MS	F^*
$E(X_1/X_2X_4)$	1	0.768	0.768	15.36
$E(X_2/X_1X_4)$	1	0.13	0.13	2.64
$E(X_4/X_1X_2)$	1	0.049	0.049	0.98← الأصغر
Error	9	0.450	0.05	

$5.12 = F(1,9,0.95) = Fout$ ← يحذف X_4

$$Y = f(X_1X_2)$$

الخطوة الثالثة:

s.o.v	d.f	SS	MS	F^*
$E(X_1/X_2)$	1	10.296	10.296	206.33
$E(X_2/X_1)$	1	0.108	0.108	2.164 ← الأصغر
Error	10	0.499	0.0499	

$4.96 = F(1,10,0.95) = F0ut$ ← يحذف X_2

$$Y = f(X_1)$$

الخطوة الرابعة:

$$F_1^* = \frac{ESS(X_1)}{RMS(X_1)} = \frac{SST - RSS(X_1)}{RSS(X_1)/11} = 189.39$$

$$4.8 = F(1,11,0.95) = Fout$$

← معنوي X_1

أفضل معادلة انحدار: $Y = \beta_0 + \beta_1 X_1 + u$

مثال (5−9): الجدول يوضح نتائج انحدار Y على توليفات مختلفة للمتغيرات X_1 ، X_2 ، X_3 لعينة بحجم (15) مشاهدة :

$$\Sigma(y-\bar{y})^2 = 1325$$

المتغيرات	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$
ESS	325	242	60	414	645	348	725

م/ حدد أفضل معادلة انحدار باستعمال الحذف العكسي . (استخدم مستوى دلالة 5%)

الخطوةالأولى :

$$F_1^* = \frac{ESS(X_1/X_2X_3)}{RSS(X_1X_2X_3)(n-4)} = \frac{725-348}{(1325-725)/11} = 6.911$$

$$F_2^* = \frac{725-645}{54.545} = 1.46668 \leftarrow$$ الأصغر

$$F_3^* = \frac{725-44}{54.545} = 5.702$$

$$F_{out}(1,11,0.95) = 4.84$$

أقل قيمة F جزئية هي F_2 وهي أقل من Fout .

لذايحذف المتغير X_2 من المعادلة

الخطوة الثانية:

$$F_1^* = \frac{ESS(X_1/X_3)}{RSS(X_1X_3)/(n-3)} = \frac{645-60}{(1325-645)/12} = \frac{585}{56.67}$$

$$F_1^* = 10.3$$

$$F_3^* = \frac{645-325}{56.67} = 5.647 \leftarrow$$ الأصغر

أصغر F جزئية هي F_3

$$Fout(1,12,0.95) = 4.75$$

$F_3^* > Fout$ ← نتوقف

إي إن أفضل معادلة انحدار هي : $Y = f(X_1, X_3)$

$$R^2 = 0.486 Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + u$$

(2−9) طريقة الاختيار المباشر أو الأمامي : (Forward Selection Procedure)

تعتمد هذه الطريقة على البدء بدون إي متغير مستقل ، ثم يتم اختيار المتغيرات المستقلة لتضمينها في المعادلة واحداً تلو الآخر ، اعتماداً على مقارنة F الجزئية لكل متغير مع قيمة جدوليه بمستوى دلالة محدد مسبقاً (α) وتسمى (F_{IN}) . إذ يتم اختيار أعلى قيمة F الجزئية لكل خطوة وبعد التأكد من إن قيمتها اكبر من (F_{IN}) يتم إدخال المتغير المعني إلى المعادلة وتستمر الخطوات بإضافة المتغيرات المستقلة واحداً تلو الآخر إلى إن نصل إلى إن أعلى (F) جزئية تقل عن (F_{IN}) . فعندئذ يتوقف الحل.

مثال (6−9) : استعمل الاختيار الأمامي مستعيناً بيانات المثال (2−9):

الخطوة الأولى:

$$F_1^* = \frac{\mathrm{ESS}(X_1)}{\mathrm{RSS}(X_1)/(n-2)} = \frac{61.474}{\frac{1388-61.474}{11}} = \frac{61.474}{120.593} = 0.51$$

$$F_2^* = \frac{404.327}{89.425} = 4.521$$

$$F_3^* = \frac{558.054}{75.449} = 7.396 \leftarrow \quad \text{الأكبرير شحوهومعنوي}$$

$$F_4^* = 151.844/112.378 = 1.351$$

$$Y = f(X_3)$$

الخطوة الثانية :

$$F_1^* = \frac{ESS(X_1/X_3)}{RSS(X_1X_3)/(n-3)} = \frac{0.964}{82.898} = 0.012$$

$$F_2^* = \frac{\mathrm{ESS}(X_2/X_3)}{\mathrm{RSS}(X_2X_3)/(n-3)} = \frac{67.461}{76.2485} = 0.884$$

$$F_4^* = \frac{\mathrm{ESS}(X_4/X_3)}{\mathrm{RSS}(X_3X_4)/(n-3)} = \frac{313.782}{51.616} = 6.079 \quad \leftarrow \quad \text{يرشح الاعلى وهو معنوي}$$

$$Y = f(X_3X_4)$$

الخطوة الثالثة:

$$F_1^* = \frac{ESS(X_1/X_3X_4)}{RSS(X_1X_3X_4)/(n-4)} = \frac{0.193}{57.33} = 0.003$$

$$F_2^* = \frac{ESS(X_2/X_3X_4)}{RSS(X_2X_3X_4)/(n-4)} = \frac{28.218}{54.216} = 0.5 \quad \leftarrow$$ الاكبر

أقل من F_{IN} نتوقف، وبذلك فان أفضل معادلة انحدار هي :

$$Y = \beta_0 + \beta_3X_3 + \beta_4X_4 + u$$

ويمكن حل المثال (9−5) اعتماداً على فكرة معاملات الارتباط ، حيث يتم اختبار المتغير المستقل الذي تكون قيمة معامل ارتباطه البسيط مع Y (r_{iy}) أعلى قيمة في كل خطوة ثم تختبر معنويته باستخدام اختبار F الجزئي .

وهناك طريقة اخرى للاختبار الامامي باستخدام بيانات مثال(9−2) نفسها:

الخطوة الأولى: باستخدام القانون:

$$r_{iY} = \sqrt{\frac{ESS(X_i)}{TSS}} \quad \forall \quad i = 1,2,3,4$$

نتوصل الى معاملات الارتباط بين المتغير Y وكل من X_1 ، X_2 ، X_3 ، X_4:

$$r_{4Y} = 0.33 \quad ، \; r_{3Y} = 0.634 \quad ، \; r_{2Y} = 0.54 \quad ، \; r_{2Y} = 0.2$$

حيث ان r_{3Y} هي اعلى معامل ارتباط

اذا نختبر معنوية X_3 وفق اختبار F:

$$F_3^* = \frac{558.054}{829.946/11} = \frac{558.054}{75.449} = 7.396$$

$F_{IN(1,11,0.95)} = 4.48$ ، أي ان X_3 معنوي

$\leftarrow$ X_3 يضمن في المعادلة : $Y = f(X_3)$

الخطوة الثانية :

$$r_{iY.3} = \sqrt{\frac{ESS(X_i/X_3)}{RSS(X_3)}} \quad \forall \quad i = 1,2,4$$

$$r_{1Y.3} = \sqrt{\frac{559.018 - 558.054}{1388 - 558.054}} = \sqrt{\frac{0.964}{829.946}}$$

$$r_{1Y.3} = 0.034$$

$$r_{2Y.3} = 0.285$$

$r_{4Y.3} = 0.615 \quad \Leftarrow$ الاعلى

يرشح X_4 $\leftarrow$

$$F_4^* = \frac{ESS(X_4/X_3)}{RSS(X_3X_4)/(n-3)} = 6.074 \quad > \quad F_{IN(1,10,0.96)} = 4.96$$

$\leftarrow$ يثبت في المعادلة : $\quad Y = f(X_3, X_4)$

الخطوة الثالثة :

$$r_{iY.34} = \sqrt{\frac{ESS(X_i/X_3X_4)}{RSS(X_3X_4)}} \qquad \forall \quad i = 1, 2$$

$$r_{1Y.34} = 0.0194$$

$r_{2Y.34} = 0.234 \quad \leftarrow$ الاعلى

يرشح X_2 $\leftarrow$

$$F_2^* = \frac{ESS(X_2/X_3X_4)}{RSS(X_2X_3X_4)/(n-4)} = \frac{900.054 - 871.836}{54.216} = 0.52$$

$$F_{IN(1,9,0.95)} = 5.12$$

وحيث ان، $F_2^* < F_{IN}$ إذن ينتهي الحل ، وبذلك فان افضل معادلة انحدار:

$$Y = f(X_3, X_4)$$

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + u$$

مثال(9−7): استخدم بيانات المثال(9−3) لاختبار افضل معادلة انحدار على وفق طريقة الاختيار الأمامي.

الخطوة الأولى:

$$r_{iY} = \sqrt{\frac{ESS(X_i)}{TSS}}$$

$$r_{1Y} = 0.6$$

$$r_{2Y} = 0.657 \quad \Leftarrow \quad \text{الاكبر}$$

$$r_{3Y} = 0.298$$

X_2 ← يرشح لتضمينه بالمعادلة

$$F_2^* = \frac{ESS(X_2)}{RSS(X_2)/(n-2)} = \frac{63}{4.6} = 13.69 \quad > \quad F_{IN(1,18,0,95)} = 4.41$$

أي ان الصيغة هي: $Y = f(X_2)$

الخطوة الثانية :

$$r_{iY.2} = \sqrt{\frac{ESS(X_i/X_2)}{RSS(X_2)}} = \sqrt{\frac{ESS(X_iX_2) - ESS(X_2)}{RSS(X_2)}}$$

$$r_{1Y.2} = \sqrt{\frac{66-63}{83}} = 0.190$$

$$r_{3Y.2} = \sqrt{\frac{79-63}{83}} = 0.4390$$

يتبين ان معامل ارتباط Y مع X_3 مع تثبيت X_2 تعد الاكبر، لذلك يرشح المتغير X_3ونختبر أهميته :

$$F_3^* = \frac{ESS(X_3/X_2)}{ESS(X_2X_3)/(n-3)} = \frac{16}{3.94} = 4.06 \quad \langle \quad F_{IN(1,16,0.95)} = 4.45$$

أذن ينتهي الحل:

أفضل معادلة انحدار هي : $Y = \beta_0 + \beta_2 X_2 + u$

مثال(9−8) اعتمد طريقةالاختبارالأمامي"Forward Selection" لاختبار أفضل معادلة انحدار اذا علمت ان مجموع مربعات الانحدار لتوافيق مختلفة من المتغيرات (X_1 ، X_2 ، X_3 ، X_4) معطاة في الجدول ادناه لعينة من (13 مشاهدة)، ومجموع المربعات الإجمالية $TSS = 2715.76$

ESS	1460.07	1809.4	776.3	1831.8	2667.8
Var	X_1	X_2	X_3	X_4	$X_1X_2X_3X_4$

↓

2641	1846.8	2540	2664.9	2667.7
X_1X_4	X_2X_4	X_3X_4	$X_1X_3X_4$	$X_1X_2X_4$

الحل:

الخطوة	في المعادلة X_i	المضاف X_i	F الجزئية	F_{IN}	المنتخب X_i
الأولى		X_1 X_2 X_3 X_4	12.8 21.9 4.4 22.8	3.23	X_4
الثانية	X_4	X_1 X_2 X_3	108.23 0.173 40.29	3.29	X_1
الثالثة	X_1X_4	X_2 X_3	5.02 4.23	3.36	X_2
الرابعة	$X_1X_2X_4$	X_3	0.02	3.46	–

(3-9) طريقة الانحدار المتدرج : Stepwise regression procedure

وتعد هذه الطريقة تطويراً لطريقة الاختيار الأمامي (Forward Selection) فقد وضع أساسها (Efroymson 1960) . لجعلها أكثر كفاءة. ونقطة التمييز بين الطريقتين هو إن جميع المتغيرات المستقلة في نهاية كل خطوة يتم التأكد من معنويتها بالاعتماد على اختيار F الجزئي ويعاد تقييمها مرة أخرى . وذلك لوجود علاقات قوية بين المتغيرات المستقلة والتي تم إدخالها في معادلة الانحدار في خطوات سابقة . وبذلك فان طريقة الانحدار المتدرج (التدريجي) تعد أفضل الطرائق لاختيار أحسن معادلة انحدار.

ولتوضيح فكرة هذه الطريقة نقدم الأمثلة التالية :

مثال (9–9): استعمل طريقة الانحدار المتسلسل " Stepwise regression" لاختبار أفضل معادلة انحدار اذا توفرت البيانات التالية: $n=10$ ، $Y'Y=396$ ، $\Sigma Y=50$

$$R=\begin{bmatrix} 1.00 & -0.988 & 0.896 & 0.972 \\ & 1.00 & -0.913 & -0.9795 \\ & & 1.00 & 0.946 \\ & & & 1.00 \end{bmatrix}$$

الحل:

الخطوة الأولى: أعلى معامل ارتباط بسيط بين المتغيرات التوضيحية وبين المتغير Y هو :

$|r_{YX_2}| = |-0.9795|$ ← المتغير X_2 هو المتغير المرشح لتضمينه في المعادلة .

نختبر معنويته باختبار F:

$$F_2^* = \frac{ESS(X_2)}{RSS(X_2)/(n-2)}$$

$$ESS(X_2) = r_{YX_2}^2 .TSS = (-0.9795)^2\,(146) = (0.9594)(146) = 140.075$$

$$RSS(X_2) = TSS - ESS(X_2) = 146 - 140.075 = 5.925$$

$$\to F_2^* = \frac{140.075}{0.7406} = 189.137 \qquad , \qquad F_{C(1\,,8,0.95)} = 5.32$$

المتغير X_2 معنوي يثبت في المعادلة :

$$Y = \beta_0 + \beta_2 X_2 + u$$

الخطوة الثانية :

$$r_{YX_1.X_2} = \left(\sqrt{\frac{ESS(X_1/X_2)}{RSS(X_2)}}\right)$$

أو

$$= \frac{r_{1Y} - r_{12}r_{2Y}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{2Y}^2}} \; \frac{0.972-(-0.988)(-0.9795)}{\sqrt{1-(-0.988)^2}\sqrt{1-(-0.9795)^2}}$$

$$=0.134$$

$r_{YX_3.X_2} = 0.62 \quad \to \quad X_3$ (included in the equation) :يضمن في المعادلة

نختبر معنوية X_3:

$$F_3^* = \frac{ESS(X_3/X_2)}{RSS(X_2X_3)/(n-3)}$$

$$r^2_{YX_3.X_2} = \frac{ESS(X_3/X_2)}{RSS(X_2)}$$

$$\Rightarrow \quad ESS(X_3/X_2) = r^2_{YX_3.X_2} RSS(X_2) = (0.3844)(5.925)$$

$$ESS(X_3/X_2) = 2.27757$$

$$ESS(X_3) = r^2_{YX_3}.TSS = 130.657736$$

$$ESS(X_2X_3) = 142.353 \quad , \quad RSS(X_2X_3) = 3.647$$

$$F_3^* = \frac{2.27757}{1.866} = 0.521 \quad , \quad F_{c(1,7,0.95)} = 5.59$$

∴ X_3 لم تثبت معنويته، فيحذف من المعادلة.

وبذلك فان أفضل معادلة انحدار هي:

$$Y = \beta_0 + \beta_2 X_2 + u$$

وللمعلومات نفسها يتم استخدام طريقة الاختبار الامامي للمقارنة وتكون خطوات الانحدار المتسلسل نفسها.

مثال(9-10): عينة من (13) مشاهدة، لانحدار Y على X_1 ، X_2 ، X_3 ، X_4 ، اعطت النتائج التالية:

$$TSS = 2715.76$$

ESS	1460.07	1809.4	776.3	1831.8	2667.8
Var	X_1	X_2	X_3	X_4	$X_1X_2X_3X_4$

↓

57.99	2641	1846.8	2540	48.11	2664.9	2667.7
X_1X_2	X_1X_4	X_2X_4	X_3X_4	$X_1X_2X_3$	$X_1X_3X_4$	$X_1X_2X_4$

م/اعتمد طريقة الانحدار التدرجي (المتسلسل) "Stepwise Regression" لاختيار أفضل معادلة انحدار.

الحل:

الخطوة	X_iفيالمعادلة	X_iالمضاف	F الجزئي	F_{IN}
الأولى		X_1 X_2 X_3 X_4	12.8 21.9 4.4 22.8 ←	3.23
الثانية	X_4 X_1	X_1 X_2 X_3 X_4	←108.23 0.173 40.29 157.963	3.29
الثالثة	X_1X_4 X_1X_2 X_2X_4	X_2 X_3 X_4 X_1	5.02 4.699 ←488.7 153.73	3.36
الرابعة	$X_1X_2X_4$	X_3	يحذف0.016←	3.36

أي ان افضل معادلة انحدار هي: $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_4X_4+u$

مثال(9−11): اعتمد طريقة الانحدار المتسلسل لاختيار أفضل معادلة انحدار ، اذا توافرت البيانات التالية.

$$[x'x \;\vdots\; x'y]=\begin{bmatrix} 74 & -10 & 3 & \vdots & 33 \\ & 14 & -1 & \vdots & 4 \\ & & 17 & \vdots & 13 \end{bmatrix} \quad , \quad \Sigma y^2=74 \quad , \quad n=25$$

$$TSS=\sum y^2=74$$

باستخدام مستوى دلالة 5%. وقرب النتائج إلى مرتبتين عشريتين فقط .

الحل :

الخطوة الأولى:

$$r_{iy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \qquad \forall \quad i = 1,2,3$$

$$r_{1y} = \frac{33}{\sqrt{(74)(74)}} = 0.45 \quad \leftarrow \quad \text{الاكبر}$$

$$r_{2y} = \frac{4}{\sqrt{(14)(74)}} = \frac{4}{32.19} = 0.12$$

$$r_{3y} = \frac{13}{35.47} = 0.37$$

المتغير X_1 يرشح، ونحسب قيمة F الجزئية .

$$F_1^* = \frac{ESS(X_1)}{RSS(X_1)/(n-2)}$$

$$ESS(X_1) = \frac{(\sum x_1 y)^2}{\sum x_1^2} = \frac{(33)^2}{74} = 14.72 \quad , \quad RSS(X_1) = 74 - 14.72 = 59.28$$

$$> \qquad F_{IN(1,23,0.95)} = 4.28 F_1^* = \frac{14.72}{2.58} = 5.71$$

المتغير X_1 يضمن في المعادلة

$$Y = f(X_1)$$

الخطوة الثانية :

$$r_{iY.1} = \sqrt{\frac{ESS(X_i/X_1)}{RSS(X_1)}} \qquad \forall \quad i = 2,3$$

$$r_{2Y.1} = \sqrt{\frac{ESS(X_2/X_1)}{RSS(X_1)}} \quad , \quad ESS(X_1 X_2) = \hat{\beta}' x' y$$

$$|x'x|_{12} = 936$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 74 & -10 \\ -10 & 14 \end{pmatrix}^{-1} \begin{pmatrix} 33 \\ 4 \end{pmatrix} = \begin{pmatrix} 0.0149 & 0.0107 \\ 0.0107 & 0.0791 \end{pmatrix} \begin{pmatrix} 33 \\ 4 \end{pmatrix} = \begin{pmatrix} 0.54 \\ 0.67 \end{pmatrix}$$

$$ESS(X_1, X_2) = \begin{pmatrix} 0.54 & 0.67 \end{pmatrix} \begin{pmatrix} 33 \\ 4 \end{pmatrix} = 20.5 \quad , \quad RSS(X_1, X_2) = 53.5$$

$$ESS(X_1) = \frac{(33)^2}{74} = 14.72$$

$$ESS(X_2/X_1) = 20.5 - 14.72 = 5.78 \quad , \quad RSS(X_1) = 59.28$$

$$r_{2Y.1} = \sqrt{\frac{ESS(X_2/X_1)}{RSS(X_1)}} = 0.3123$$

$$r_{3Y.1} = \sqrt{\frac{ESS(X_3/X_1)}{RSS(X_1)}}$$

$$ESS(X_1X_3) = (\hat{\beta}_1 \quad \hat{\beta}_3)\begin{pmatrix}33\\13\end{pmatrix}$$

$$\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{bmatrix}74 & 3\\ & 17\end{bmatrix}^{-1}\begin{pmatrix}33\\13\end{pmatrix}$$

$$|x'x|_{1.3} = 1249 \quad \Rightarrow \quad (x'x)^{-1}_{1.3} = \begin{bmatrix}0.014 & -0.0024\\ & 0.06\end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_3\end{bmatrix} = \begin{pmatrix}0.43\\0.70\end{pmatrix}$$

$$ESS(X_1X_3) = 23.29$$

$$ESS(X_3/X_1) = 23.29 - 14.72 = 8.57$$

$$r_{3Y.1} = \sqrt{\frac{8.57}{59.28}} = 0.380$$

الأكبر، $r_{3Y.1} = 0.380$ ← المتغير X_3 يرشح

$$F_3^* = \frac{ESS(X_3/X_1)}{RSS(X_1X_3)/(n-3)} = \frac{8.57}{23.29/22} = \frac{8.57}{1.059} = 8.093 \quad \rangle \quad F_{IN(1,22,0.95)} = 4.3$$

$$F_1^* = \frac{ESS(X_1/X_3)}{RSS(X_1X_3)/(n-3)} == \frac{13.349}{1.059} = 12.605 \quad \rangle \quad F_{IN(1,22,0.95)} = 4.3$$

حيث إن :

$$ESS(X_1/X_3) = ESS(X_1X_3) - ESS(X_3) = 23.29 - \frac{13^2}{17} = 13.349 \quad \rangle \quad F_{IN} = 4.3$$

النموذج . $Y = f(X_1, X_3)$

$$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + u$$

الخطوة الثالثة:

$$r_{2Y.1,3} = \sqrt{\frac{ESS(X_2/X_1X_3)}{RSS(X_1X_3)}}$$

$$ESS(X_2/X_1X_3) = ESS(X_1X_2X_3) - ESS(X_1X_3)$$

$$ESS(X_1X_2X_3) = \left[(x'x)^{-1}x'y\right] x'y$$

مثال(12−9): اذا توفرت البيانات التالية:

المتغيرات	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3
RSS	1225.68	906.33	1939.01	883.86	57.90	1227.07	74.76	415.44

X_2X_4	X_3X_4	$X_1X_2X_3$	$X_1X_3X_4$	$X_2X_3X_4$	$X_1X_2X_4$	$X_1X_2X_3X_4$
868.88	175.75	48.11	50.83	73.81	47.97	47.86

$TSS = 2715.63 \quad , \quad n = 13$

استخدم طريقة الانحدار التدريجي لإيجاد أفضل معادلة انحدار. باستخدام مستوى دلالة 5% .

الحل:

الخطوة الأولى:

$$r_{1Y} = \sqrt{\frac{ESS(X_1)}{TSS}} = \sqrt{\frac{1489.95}{2715.63}} = 0.74$$

$r_{2Y} = 0.816$

$r_{3Y} = 0.534$

$r_{4Y} = 0.8213 \quad \leftarrow$ أكبر

نختار X_4 أولاً .

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_4 X_4$$

ونستخدم المختبر F لاختبار معنوية X_4:

$$F_4^* = \frac{ESS(X_4)}{RMS(X_4)} = \frac{1831.77}{80.35} = 22.797 \quad , \quad F_{c(1,11,0.95)} = 4.4 \quad \text{الجدولية}$$

$\leftarrow$ يثبت X_4 في النموذج.

الخطوة الثانية :

$$r_{1Y.4} = \sqrt{\frac{ESS(X_1/X_4)}{TSS - ESS(X_4)}} = \sqrt{\frac{ESS(X_1X_4) - ESS(X_4)}{RSS(X_4)}} = 0.957$$

$r_{2Y.4} = 0.1302$

$r_{3Y.4} = 0.895$

لذا ينتخب X_1، فتكون معادلة الانحدار: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_4 X_4$

ونختبر معنوية المتغير X_4

$$F_4^* = \frac{ESS(X_4/X_1)}{RMS(X_1, X_4)} = 153.9 \quad \rangle \quad F_{c(1,n-3,0.05)}$$

$$F_1^* = \frac{ESS(X_1/X_4)}{RMS(X_1, X_4)} = 108.2 \quad \rangle \quad F_{c(1,n-3,0.09)}$$

← يثبت كل من X_1 و X_4 في النموذج.

الخطوة الثالثة:

$$r_{2Y.14} = \sqrt{\frac{ESS(X_2/X_1X_4)}{TSS - ESS(X_1X_4)}}$$

$$r_{2Y.14} = 0.599$$

$$r_{3Y.14} = 0.565$$

لذا نختار X_2 ويضمن في المعادلة:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X_4$$

وتحسب معنوية X_2 المضاف:

$$F_2^* = \frac{ESS(X_2/X_1X_4)}{RMS(X_1X_2X_4)} = \frac{26.8}{5.33} = 5.02 \quad \rangle \quad F_{c(1,n-4,0.95)}$$

$$F_1^* = \frac{ESS(X_1/X_2X_4)}{RMS(X_1X_2X_4)} = \frac{820.9}{5.33} = 154.01 \quad \rangle \quad F_{c(1,n-4,0.95)}$$

ولكن F_4^* مع وجود X_2 و X_1: $F_4^* = 1.86$ غير معنوي (لذا يحذف X_4 من النموذج) .

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

الخطوة الرابعة:

يضاف المتغير X_3

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

وتحسب معنوية X_3:

$$F_3^* = \frac{ESS(X_3/X_1X_2)}{RMS(X_1X_2X_3)} = \frac{9.79}{5.35} = 1.83 \quad \rangle \quad F_{c(1,n-4,0.95)}$$

وبمقارنتها مع القيم الجدولية يتضح ان X_3 غير معنوي.

لذا يحذف X_3.

وتكونأفضلمعادلةانحدار : $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$

كماإن مثال(9−10) يمكن حله بطريقة مغايرة اعتماداًعلى قيمF الجزئية.

مثال (9−13) باستخدام F

الخطوة الأولى :

$$F_1^* = \frac{ESS(X_1)}{RMS(X_1)} = \frac{TSS - RSS(X_1)}{RSS(X_1)/(n-k-1)}$$

$$F_1^* = 13.37$$

$$F_2^* = 21.96$$

$$F_3^* = 4.41$$

$$F_4^* = 22.797 \qquad , \quad F_{IN} = F_{c(1,11,0.95)} = 4.48$$

← X_4 يدخل المعادلة

الخطوة الثانية:

$$F_1^*(X_1/X_4) = 108.2$$

$$F_2^*(X_2/X_4) = 0.172$$

$$F_3^* = 40.291 \qquad , \quad F_{IN} = F_{c(1,10,0.95)} = 4.96$$

← المتغير X_1 يدخل المعادلة

وتحسب معنوية X_4 بعد ادخال X_1 المعادلة

$$F_4^*(X_4/X_1) = 153.9$$

الخطوة الثالثة:

$$F^*_{2(X_2/X_1X_4)}) = 5.02 \qquad F^*_{3(X_3/X_1X_4)}) = 4.23 \quad , \quad F_{IN} = 5.12$$

نتوقف: افضل معادلة انحدار $\Leftarrow \quad Y = \beta_0 + \beta_4 X_4 + u$

مثال(9−14): من لوحة المعلومات التالية :

$TSS = 146 \qquad , \qquad n = 20$

RSS	93	83	133	80	76	67	65
المتغيرات	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$

حددأفضل معادلة انحدار باستخدام الاختيار الأمامي وقارنها مع الانحدار التدرجي .

القيم الجدولية :

$F(10,5,0.95) = 4.74 \quad , F(1,18,0.95) = 4.41 \quad , t(37,0.025) = 2.028$

$t(16,0.025) = 2.120 \quad , F(1,17,0.95) = 3.59 \quad ,F(1,16,0.95) = 3.24$

الحل: باستخدام الاختبار الامامي:

الخطوة الأولى:

$F_1^* = 10.258$

$F_2^* = 13.66$ ⇐ الاكبر $F_{C(1,18)} = 4.4$ ⇐ X_2 يضمن

$F_3^* = 1.759$

$Y = f(X_2)$

الخطوة الثانية:

$F_{1.2}^* = 0.638$

$F_{3.2}^* = 4.059$ ⇐ الاكبر $F_{c(1,17)} = 3.59$ ⇐ X_3 يضمن

$$Y = f(X_2, X_3)$$

الخطوة الثالثة:

$$F_{1.23} = \frac{ESS(X_1/X_2, X_3)}{RSS(X_1X_2, X_3)/(n-4)}$$

$$= \frac{(146-65)-(146-67)}{65/(20-4)} = \frac{2}{4.0625} = 0.49$$

F غير معنوية نتوقف

أفضل معادلة انحدار: $Y = \beta_0 + \beta_2X_2 + \beta_3X_3 + u$

الانحدار المتسلسل : الخطوة الأولى : نفسها كما في الاختبار الأمامي.

الخطوة الثانية:أيضا تحسب

$$F_{2.3} = \frac{79-13}{67/17} = \frac{66}{3.94} = 16.75$$

يثبت X_2

الخطوة الثالثة: نفسها

مثال(15–9) : مع توافر لوحة البيانات:

$$x'x = \begin{bmatrix} 74 & 10 & 3 \\ & 14 & 1 \\ & & 17 \end{bmatrix} \; ; \; x'y = \begin{bmatrix} 33 \\ 4 \\ 13 \end{bmatrix} ; \; \sum y^2 = 64 \; ; n = 25$$

استخدم طريقة الانحدار التسلسلي لتقدير أفضل معادلة انحدار لاستجابة Y والمتغيرات X_1, X_2, X_3 استخدم مستوى دلالة 10%، اذكر الخطوتين الأولى والثانية فقط.

الحل:

الخطوة الأولى:

$$r_{1Y} = \frac{\Sigma x_1 y}{\sqrt{\Sigma x_1^2 \Sigma y^2}} = \frac{33}{\sqrt{(74)(64)}} = 0.479$$

أو يمكن حسابها على وفق الاتي:

$$= \sqrt{\frac{ESS(X_1)}{TSS}} = \sqrt{\frac{14.68}{64}} = 0.479 \leftarrow$$ الاكبر

وكذلك:

$$r_{2Y} = \frac{\sum x_2 y}{\sqrt{\sum x_2^2 \sum y^2}} = 0.133$$

$$r_{3Y} = \frac{\sum x_3 y}{\sqrt{\sum x_3^2 \sum y^2}} = 0.39$$

وحيث ان معامل ارتباط Y مع X_1 هو الاكبر، لذا نختار المتغير X_1 لتضمينه في المعادلة، ثم نحسب قيمة F لاختبار معنوية المتغير X_1

$$F_1^* = \frac{ESS(X_1)}{RMS(X_1)} = \frac{r_{1Y}^2 \sum y^2}{RSS(X_1)/23} = \frac{14.68}{2.144} = 6.847$$

وحيث ان قيمة F الجدولية هي: $(F_{1N(1,23,0.95)} = 2.94)$

X_1 يضمن في المعادلة:

- X_1 مهم يثبت في المعادلة.

الخطوة الثانية: نحسب معاملات الارتباط الجزئية بين المتغير Y والمتغيرات المستقلة X_2 و X_3 علماً بان X_1 ثابت

$$r_{2Y.1} = \frac{r_{2Y} - r_{12}\, r_{1Y}}{\sqrt{(1-r_{12}^2)(1-r_{1Y}^2)}}$$

وذلك يتطلب حساب معاملات الارتباط البسيط r_{12}:

$$r_{12} = \frac{\sum x_1 x_2}{\sqrt{\sum x_1^2 \sum x_2^2}} = \frac{10}{\sqrt{(74)(14)}} = \frac{10}{32.187} = 0.311$$

$$r_{2Y.1} = \frac{0.133 - 0.311\ (0.479)}{\sqrt{1-\ (0.311)^2}\sqrt{1-(0.479)^2}} = \frac{-0.016}{0.95(0.877)} = \frac{-0.016}{0.833} = -0.019$$

وكذلك: $r_{3Y.1} = \dfrac{r_{3Y} - r_{31}\, r_{Y1}}{\sqrt{1-r_{31}^2}\sqrt{1-r_{Y1}^2}}$

والذي يتطلب حساب r_{13} :

$$r_{13} = \frac{3}{\sqrt{(74)(17)}} = 0.08$$

$$\Rightarrow \quad r_{3Y.1} = \frac{0.39 - 0.08\ (0.479)}{(0.996)\ (0.8778)} = \frac{0.352}{0.874} = 0.4$$

وبما ان $r_{3Y.1}$ هي الأكبر X_3 ← يضمن في المعادلة وتحسب معنويته على وفق اختبار F :

$$F_3^* = \frac{ESS\ (X_3|X_1)}{RSS(X_1X_3)|(n-3)}$$

ولحساب $ESS(X_3/X_1)$

$$r_{3Y.1}^2 = \frac{ESS(X_3/X_1)}{RSS(X_1)} \quad \rightarrow \quad ESS(X_3/X_1) = RSS(X_1)\cdot r_{3Y.1}^2 = 0.16\ \ RSS(X_1)$$

ولكن

$$\frac{RSS(X_1)}{TSS} = 1 - r_{1Y.}^2 \quad \rightarrow \quad RSS(X_1) = \left[1-(0.479)^2\right]\ (64) = 49.32$$

$$ESS(X_3/X_1) = 0.16\ \ (49.3) = 7.888$$

$$RSS(X_1X_3) = TSS - ESS(X_1X_3)$$

$$ESS(X_1X_3) = ESS(X_1) + ESS(X_3|X_1)$$

$$= TSS - RSS + ESS(X_3/X_1)$$

$$= (14.68) + 7.888 \quad = 22.568$$

$$\Rightarrow RSS(X_1X_3) = 64 - 22.568 = 41.432$$

$$F_3^* = \frac{7.888}{41.432/(n-3)} = 4.188$$

وتقارن مع $F_{c(1,22,0.10)} = 2.95$

وبذلك نستدل ان X_3 مهم معنوياً ويضمن بالنموذج.

كما ويجب ان نستدل حول معنوية المتغير X_1 عند تضمين المتغير X_3 في النموذج.

ونختبره باستخدام F الجزئي:

$$F_1^* = \frac{ESS(X_1/X_3)}{RSS(X_1X_3)/(n-3)}$$

$$ESS(X_1/X_3) = r_{1Y.3}^2 \quad RSS(X_3) = ESS(X_1X_3) - ESS(X_3)$$

$$\rightarrow ESS(X_3) = 9.73 \quad , \quad RSS(X_3) = 54.26$$

$$ESS(X_1/X_3) = 22.568 - 9.73 = 12.838$$

$$F_1^* = \frac{12.838}{1.883} = 6.817 \quad > \quad F_{c(1,22,0.10)} = 2.95$$

إذن X_1 مهم يبقى في النموذج

$$Y = \beta_0 + \beta_1X_1 + \beta_3X_3 + u$$

أسئلة الفصل التاسع:

س١: عينة عشوائية بحجم (20)مشاهدة لقيم المتغير Y مع المتغيرات التوضيحية X_1 ، X_2 ، X_3 ، فاذا توافرت المعلومات التالية:

$$R = \begin{pmatrix} 1.00 & -0.998 & 0.851 & 0.972 \\ & 1.00 & -0.913 & -0.979 \\ & & 1.00 & 0.949 \\ & & & 1.00 \end{pmatrix} \quad , \quad \begin{matrix} \Sigma Y = 50 \\ \Sigma Y^2 = 369 \end{matrix}$$

م/ ايجاد افضل معادلة انحدار خطي باستخدام طريقة الاختيار الامامي ومقارنتها مع طريقة الانحدار التدرجي.

س2: مع توافر لوحة البيانات:

$$x'x = \begin{pmatrix} 74 & 10 & 3 \\ & 14 & 1 \\ & & 17 \end{pmatrix} \quad , \quad x'y = \begin{pmatrix} 33 \\ 4 \\ 13 \end{pmatrix} \quad , \quad \Sigma y^2 = 64 \quad , \quad n = 25$$

أ) استخدم طريقة الانحدار التسلسلي لتقدير افضل معادلة انحدار لاستجابة Y والمتغيرات X_3 ، X_2 ، X_1.

ب) قارن النتائج باستخدام طريقة الاختيار الامامي.

س3: حدد افضل معادلة انحدار باستخدام اسلوب الحذف العكسي مع توافر لوحة المعلومات حول استجابة Y على المتغيرات X_3 ، X_2 ،X_1.

المتغيرات	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$
RSS	93	83	133	80	76	67	65

س٤: مع توافر لوحة المعلومات التالية:

s.o.v	$R(X_1 X_2 X_3/X_0)$	$R(X_1 X_2/X_0)$	$R(X_1 X_3/X_0)$	$R(X_2 X_3/X_0)$	$R(X_1/X_0)$
SS	725	414	645	348	325

↓

s.o.v	$R(X_2/X_0)$	$R(X_3/X_0)$	Total
SS	242	60	1325

استنتج افضل معادلة انحدار على وفق طريقة الانحدار التدرجي.

الفصل العاشر

المتغيرات الوهمية (الصورية) Dummy Variables

في الفصول السابقة من الكتاب تمت دراسة الانحدار بالنسبة للمتغيرات التي يمكن توصيفها كمياً، غير أن هناك متغيرات أخرى لايمكن توصيفها كمياً والتي تؤثر في الظواهر الاقتصادية نوعياً مثل الجنس، الديانة ، الجنسية ،الحروب ، الزلازل أو الاضطرابات أو التغيرات في السياسات الحكومية وما الى ذلك. ويشار الى هذه المتغيرات بالمتغيرات الوهمية، أو المتغيرات الثنائية (Binary) أو المتغيرات النوعية (Qualitative variables) أو المتغيرات الفئوية (categorical) وسيخصص هذا الفصل في دراسة هذا النوع من المتغيرات.

(1-10)طبيعة المتغيرات الوهمية وكيفية استخدامها في الانحدار.

(1-1-10)طبيعة المتغيرات الوهمية:

ان المتغيرات الوهمية هي متغيرات نوعية تشير الى وجود أو عدم وجود ظاهرة معينة. وتعد أداة قوية وفاعلة لتمثيل السلوك الوصفي للمشاهدات (الأفراد). فالمتغير الوهمي يصف الحوادث والتي تمتلك صفتين فقط ، وتأخذ هذه المتغيرات قيمتين تحكميتين فقط هما الصفر والواحد(*) .

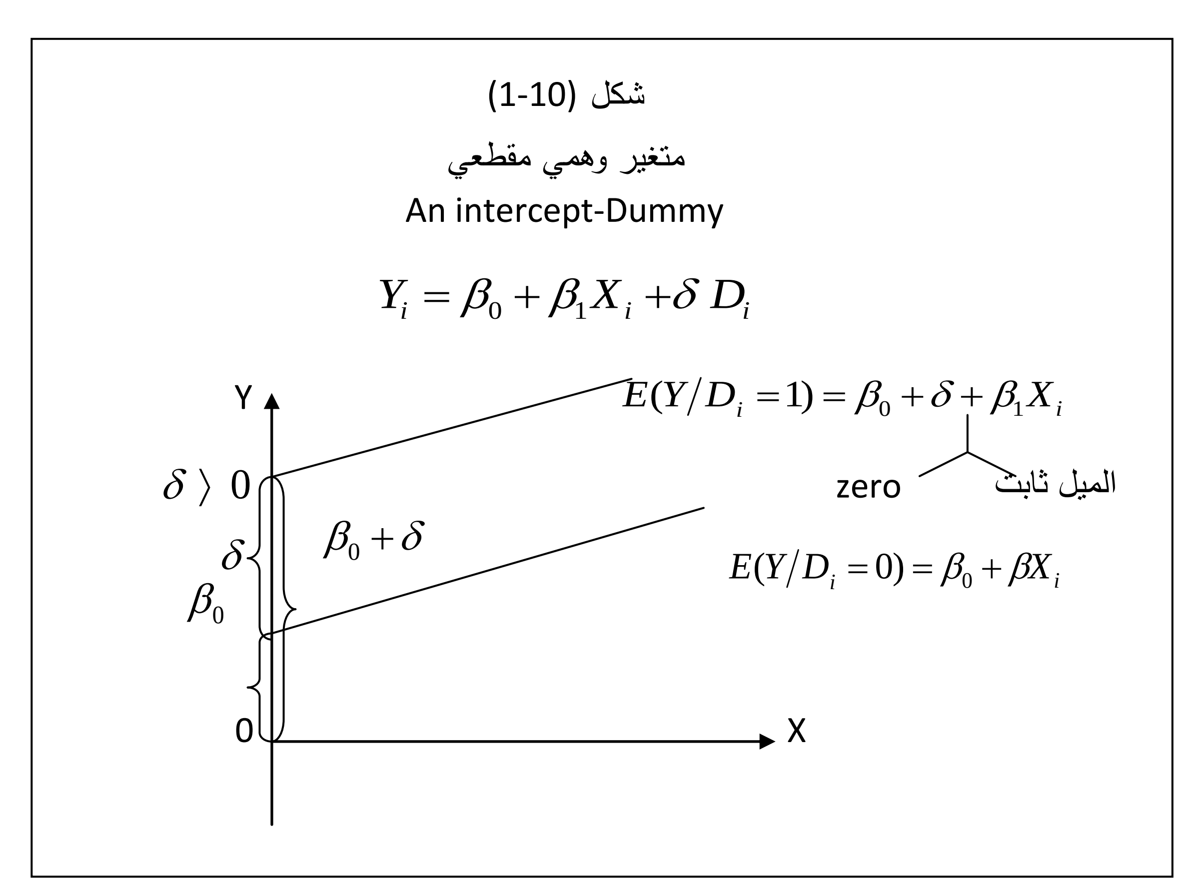
فهي تأخذ القيمة واحد(1) عند توافر الصفة وتحققها، في حين تأخذ صفر(0) عند غياب الصفة وعدم تحققها. وتكون المتغيرات الوهمية متغيرات مستقلة (التوضيحية)(وتدعى نماذج تحليل التباين (ANOVA). والتي يظهر تأثيرها فى المقطع الثابت وتسمى أيضاً (intercept-dummy-variables) وشكلها العام:

$$Y_i = \beta_0 + \beta_1 X_i + \delta\, D_i + u_i \qquad \ldots \qquad (1\text{-}10)$$

والشكل (1) يوضح ذلك.

(*) ان اختيار قيمة (0) أو (1) تكون اختيارية بحتة، حيث ان:

$$D_i = \begin{cases} 1 & \text{المشاهدة (i) تحمل الصفة} \\ 0 & \text{غير ذلك} \end{cases}$$



أما اذا احتوت النماذج على متغيرات توضيحية كمية الى جانب متغيرات توضيحية نوعية (وهمية) فتسمى نماذج تحليل التغاير(ANCOVA)، وبذلك يكون تأثيرها فى الميل ويسمى (slope dummy varibles) أو معامل التداخل من خلال ضرب المتغير الوهمي (D_i) بالمتغير الكمي (X_i)، المتغير (D_i X_i)، ويتم استخدامه عندما يكون التغير في (Y_i) نسبة الى المتغير التوضيحي (X_i) يعتمد على مستوى متغير توضيحي آخر هو D_i.

ويمكن ان توضح صيغتها العامة على وفق الآتي:

$$Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \gamma X_i D_i + u_i \qquad \ldots \qquad (2\text{-}10)$$

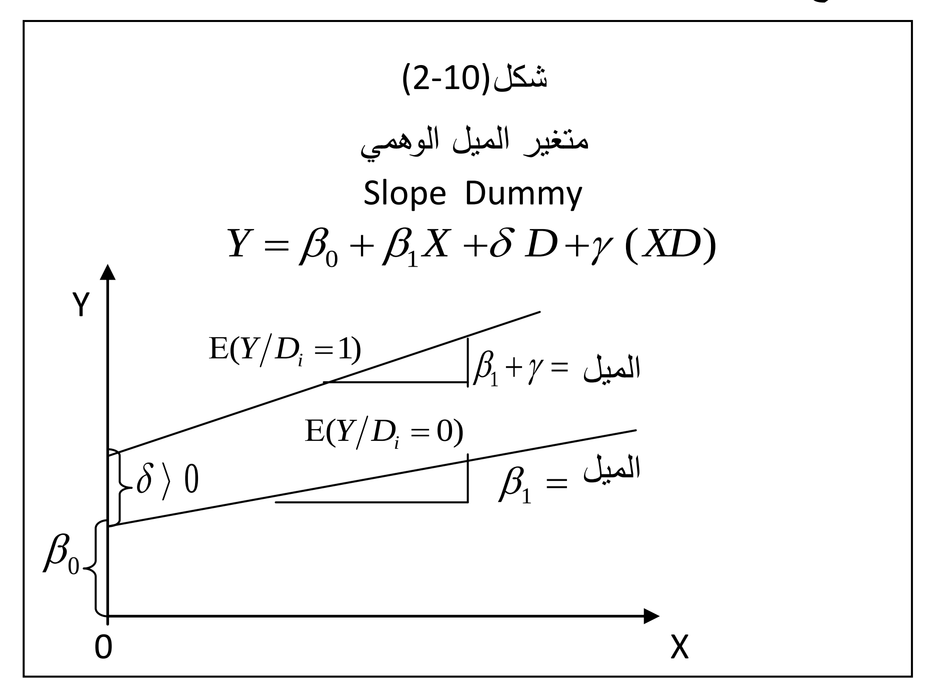
وبذلك فان الميل للدالة يمكن تحديده على وفق الاتي:

$$\frac{dY_i}{dX_i} = \beta_1 \quad , \quad D_i = 0 \quad : \text{عندما}$$

$$\frac{dY_i}{dX_i} = \beta_1 + \gamma , \quad D_i = 1$$

ولقد تم تضمين(δD_i) من اجل منع التحييز في معلمة حد الميل الوهمي ($D_i X_i$) المقدرة $(\hat{\gamma})$.

والشكل (2-10) يوضح ذلك :



حيث ان : δ : تمثل الاختلاف في الحد الثابت.

و γ : تمثل الاختلاف في الميل.

كما ان المتغيرات الوهمية قد تكون متغيرات معتمدة (Y) (dependent dummy varibles)[*] .

وجدير بالذكر ان المتغيرات الوهمية تستخدم في كلتا الدراسات المقطعية والسلاسل الزمنية.

(2-1-10) استخدامات المتغيرات الوهمية:

المتغيرات الوهمية قد تكون توضيحية أو معتمدة، وقد يكون تأثيرها فى المقطع الثابت فقط فتسمى النماذج (ANOVA) أو (intercept dummy) أو يكون تأثيرها فى الميل فتسمى (ANOCVA) أو(slope dummy) وذلك من خلال تداخلها مع المتغيرات التوضيحية الكمية وفي كلا الحالتين يمكن ان يكون تطبيقها على السلاسل الزمنية أو المقاطع العرضية أو تعمل على دمج السلاسل الزمنية بالمقاطع العرضية. وسيتم توضيح الاستخدامات من خلال الأمثلة التطبيقية التالية:

استخدامات المتغيرات الوهمية في الانحدار:

مثال(1-10): دراسات المقاطع العرضية: دراسة اسعار السكن (Y) بدلالة حجم السكن (X)، تم اختيار عينة من محافظة البصرة وبذلك فان سعر السكن Yبدلالة حجم السكن X وبالصيغة الخطية:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

β_1: تمثل قيمة المتر المربع الإضافي لوحدة السكن.

(*) موضوع المتغيرات الوهمية الداخلية هو خارج نطاق الكتاب.

β_0: تمثل قيمة الأرض فقط.

معلوم ان أسعار السكن في المحافظة تتأثر أيضاً بموقع السكن، فوجود السكن في موقع متكامل الخدمات يختلف عن سعر السكن بالحجم نفسه وفي موقع آخر أقل تكاملاً للخدمات، ويمكن مواجهة هذه المشكلة بتقدير معادلتين منفصلتين لسعر السكن وذلك عن طريق استخدام بيانات أسعار السكن في الموقع متكامل الخدمات لتقدير سعر السكن في ذلك الموقع. واستخدام البيانات لأسعار السكن في المواقع الأخرى وهكذا يتم الحصول على معادلتين مختلفتين لأسعار السكن بدلالة حجم السكن.

ولكن هناك طريقة أكثر كفاءة وهي تقدير معادلة واحدة وللمواقع جميعها ولكن بعد وضع الافتراضات الخاصة.

ولنفترض ان القيود على اسعار السكن في المواقع غير متكاملة الخدمات لاتغير الميل الحدي للسعر ولكنها تخفض الميل المتوسط له كما في الشكل (3-10).

وهكذا يمكن تضمين متغير الموقع D الى معادلة الانحدار ويتم توصيف المتغير (الموقع) كالاتي:

$$D_i = \begin{cases} 1 & \text{السكن في موقع متكامل الخدمات} \\ 0 & \text{غير ذلك} \end{cases}$$ وعلية

تصبح معادلة الانحدار:

$$Y_i = \beta_0 + \beta_1 X_i + \delta\, D_i + e_i \qquad , \qquad i = 1, \cdots, n$$

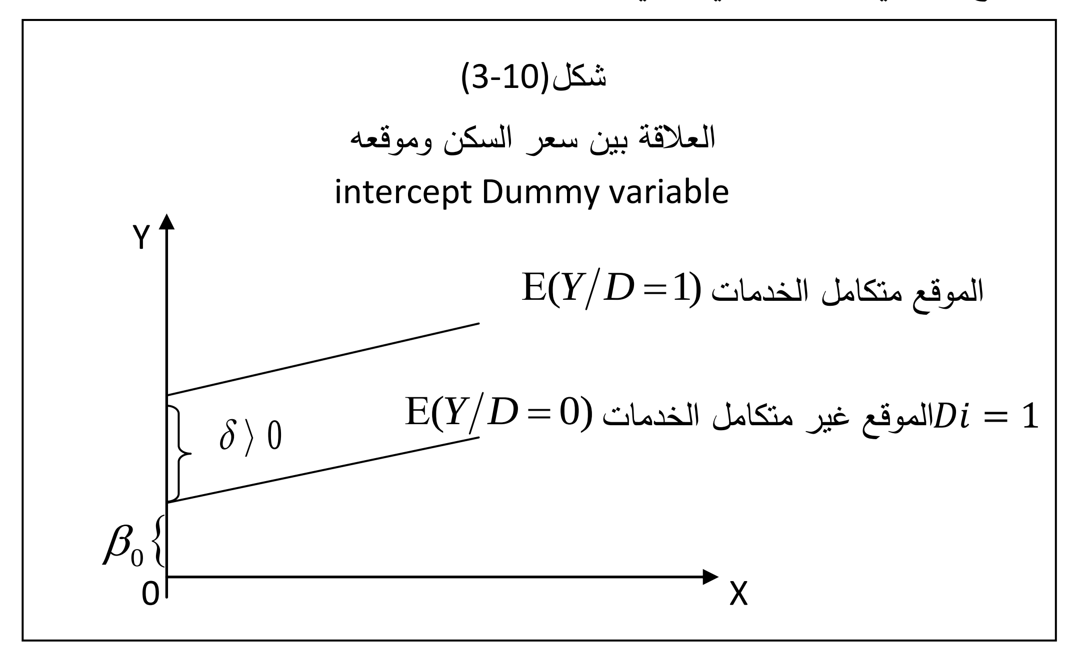
من ذلك يتبين ان متوسط الاستجابة لسعر السكن في محافظه البصرة كالآتي:

سعر السكن في الموقع متكامل الخدمات $$\mathrm{E}(Y_i / D = 1) = (\beta_0 + \delta) + \beta_1 X_i$$

سعر السكن في المواقع الأخرى $$\mathrm{E}(Y_i / D = 0) = \beta_0 + \beta_1 X_i$$

$$Y = \beta_0 + \beta_1 X + \delta\, D + \gamma\, (XD)$$

وبمكن توضيح ذلك في الشكل البياني التالي:



يتضح من الشكل (3-10) ان اضافة المتغير الوهمي D_i الى معادلة الانحدار سوف يخلق انتقالاً بشكل مواز في علاقة الانحدار وبمقدار (δ) معامل المتغير الوهمي D_i. حيث ان δ يمكن تفسيرها بانها الفرق في سعر السكن (location premium) بسبب وجوده في الموقع ذي الخدمات المتكاملة.

وفي هذه الحالة فان المتغير الوهمي D يسمى (intercept dummy variable).

أن المعلمة δ تعكس أهمية المتغير الوهمي (الموقع) وبذلك فهي تعكس اختلاف سعر السكن اعتماداً على الموقع.

وفي حال افتراض تداخل بين حجم السكن (X) وبين تكامل خدمات الموقع (D)، فيتم توصيف متغير التداخل بحاصل ضرب ($X_i D_i$) ويسمى متغير التداخل (interaction) أو قد يسمى (Slops Dummy Variable) اذ ان هذا المتغير يسمح بتغييرات في الميل لعلاقة الانحدار :

$$X_i D_i = \begin{cases} X_i & \text{الموقع متكامل الخدمات} \\ 0 & \text{الموقع غير متكامل الخدمات} \end{cases}$$

وبذلك فان متوسط الاستجابة للموقعين :

$$\mathrm{E}(Y_i) = \beta_0 + \beta_1 X_i + \gamma(X_i D_i) = \begin{cases} \beta_0 + (\beta_1 + \gamma) X_i & D_i = 1 \\ \beta_0 + \beta_1 X_i & D_i = 0 \end{cases} \quad . . . \quad (3\text{-}10)$$

بمعنى ان سعر المتر المربع للسكن $(\beta_1+\gamma)$ اذا كان موقع السكن متكامل الخدمات في حين سعر المتر المربع للسكن هو (β_1) في المواقع الاخرى .

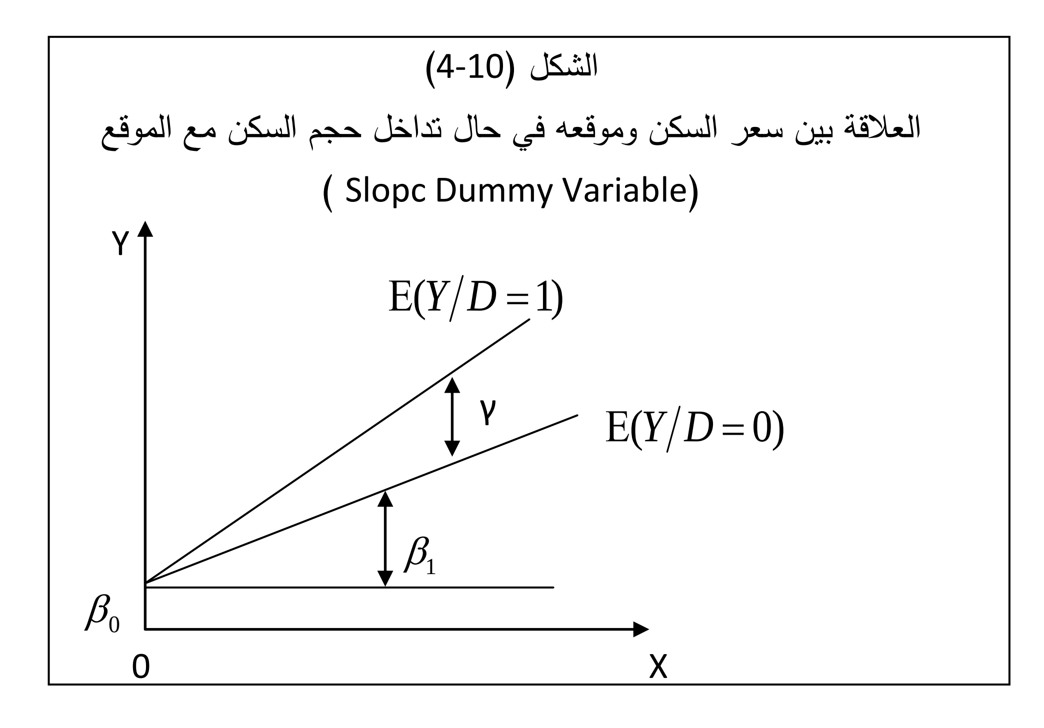
وبذلك فان γ (معامل التداخل) هو الفرق بسعر المتر المربع للسكن في الموقعين ويكون موجباً اذا كان الموقع له أفضلية على المواقع الأخرى.

والشكل (4-10) يوضح ذلك

الشكل (4-10)

العلاقة بين سعر السكن وموقعه في حال تداخل حجم السكن مع الموقع

(Slopc Dummy Variable)



وبعبارة أخرى فان اثر التداخل يمكن توضيحه من خلال المشتقة الجزئية لتوقع سعر السكن نسبة الى حجم السكن والذي يتمثل بالميل :

$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \begin{cases} \beta_1 + \gamma & D_i = 1 \\ \beta_1 & D_i = 0 \end{cases}$$

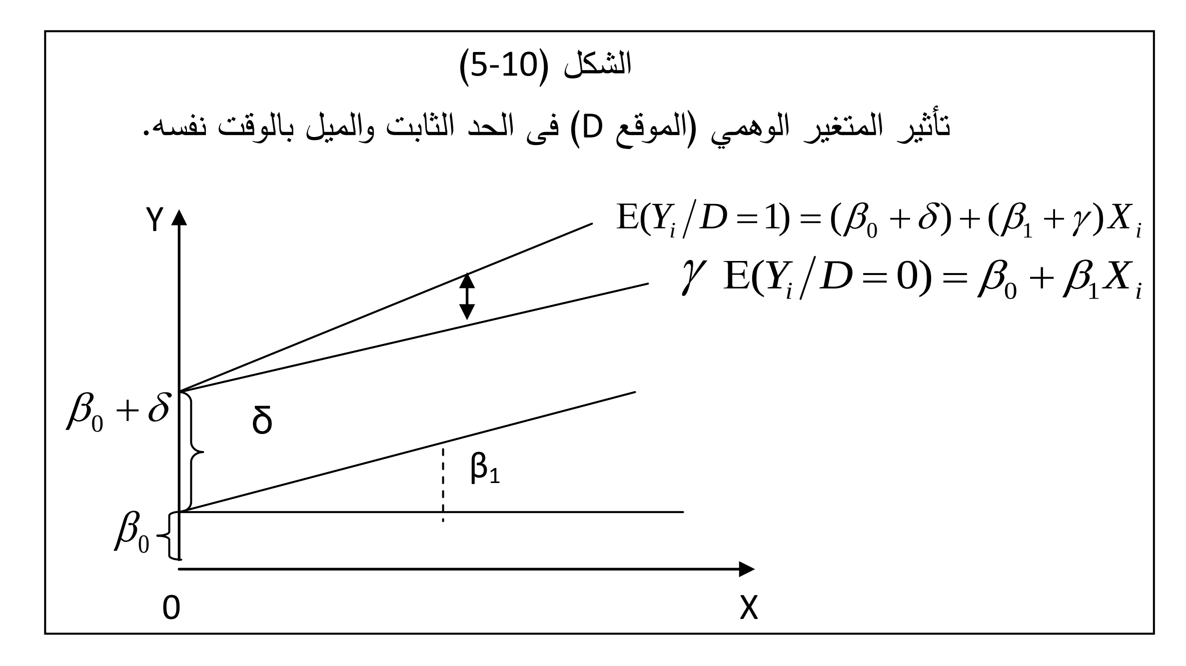
وبشكل عام يمكن افتراض ان الموقع (D) يؤثر فى المقطع والميل بالوقت ذاته فتكون معادلة الانحدار:

$$Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \gamma (X_i D_i) + e_i \qquad ... \qquad (4\text{-}10)$$

وبذلك فان متوسط سعر السكن :

$$\mathrm{E}(Y_i) = \begin{cases} (\beta_0 + \delta) + (\beta_1 + \gamma) X_i & if \quad D_i = 1 \\ \beta_0 \qquad + \ \beta_1 X_i & if \quad D_i = 0 \end{cases} \qquad ... \qquad (5\text{-}10)$$

ويمكن تمثيلها بالشكل (5-10)



الشكل (5-10)

تأثير المتغير الوهمي (الموقع D) فى الحد الثابت والميل بالوقت نفسه.

مثال (2-10): دراسة اختلاف البيئة (ريف . حضر): مقارنة دالتي الاستهلاك في الريف والحضر وذلك من خلال عينة من اسر الريف حجمها (n_1) ونقدر دالة الاستهلاك منها، وتأخذ عينة من اسر الحضر حجمها(n_2) ثم نقدر دالة الاستهلاك منها كالآتي .

دالة المستهلك في الريف : $$Y = \alpha_1 + \alpha_2 X_t \qquad ... \qquad (6\text{-}10)$$

دالة المستهلك في الحضر: $$Y = \beta_1 + \beta_2 X_t \qquad ... \qquad (7\text{-}10)$$

وبذلك فان دالة الاستهلاك لها أربعة احتمالات ممكنة :

أ) اختلاف حد الكفاف بين الحضر والريف $\alpha_1 \neq \beta_1$ وبيانيا يمكن تمثيلها بالشكل (3-10).

ب) اختلاف الميل الحدي بين الحضر والريف $\alpha_2 \neq \beta_2$ وبيانيا يمكن تمثيلها بالشكل (4-10).

ج) اختلاف بين حدي الكفاف والميل الحدي في الحضر والريف $\alpha_1 \neq \beta_1 \& \alpha_2 \neq \beta_2$ وبيانيا يمكن تمثيلها بالشكل (5-10).

د) دالتا الاستهلاك متماثلتان تماما $\alpha_1 = \beta_1 \& \alpha_2 = \beta_2$: لا اختلاف بين دالة الاستهلاك في الحضر والريف.

ويمكن اعتماد هذه الافتراضات جميعها بإدخال المتغير الوهمي، اذ يساعد استخدام المتغيرات الوهمية على اختبار كل الاحتمالات السابقة من خلال تقدير معادلة انحدار واحدة بدلاً من تقدير معادلتين ومقارنتهما. ويتم استخدام عينة كبيرة حجمها $n = n_1 + n_2$ وتكون صيغة الانحدار

$$Y_t = \alpha + \beta X_t + \delta D_t + \gamma D_t X_t + u_t \qquad \ldots \qquad (8\text{-}10)$$

$$D_i = \begin{cases} 1 & \text{الحضر} \\ 0 & \text{الريف} \end{cases}$$

ويتضح ان القيمة المتوقعة للانفاق الاستهلاكي في الريف: $E(Y/D=0) = \alpha + \beta X$

في حين القيمة المتوقعة للإنفاق الاستهلاكي في الحضر: $E(Y/D=1) = (\alpha + \delta) + (\beta + \gamma)X$

مثال (3-10): في دراسات السلاسل الزمنية: دراسة الإنفاق الاستهلاكي على سلعة معينة يعتمد على مستوى الدخل المتاح، وعند اخذ عينة لسنوات معينة منها سنوات حرب وأخرى سنوات سلم، ولإدخال تأثير وقت السلم أو وقت الحرب في معادلة الانحدار يكون بأحد الطريقتين. الأولى عزل سنوات الحرب عن سنوات السلم وتقدير دالة الاستهلاك كدالة بدلالة مستوى الدخل المتاح عن طريق استخدام بيانات منفصلة لكل فقره بشكل منعزل وبذلك يتم الحصول على معادلتين مختلفتين للاستهلاك. اما الطريقة الثانية وهي أكثر كفاءة وذلك تقدير معادلة واحدة وللفترتين بعد وضع فروض وهي ان القيود على الاستهلاك وقت الحرب لاتغير الميل الحدي للاستهلاك، ولكنها تخفض الميل المتوسط له وبذلك نفترض ان دالة الاستهلاك خلال سنوات الحرب لها الميل نفسه في سنوات السلم ولكن لها حد ثابت اصغر وبذلك فان دالة الاستهلاك لإجمالي المدة يمكن التعبير عنها بمعادلة انحدار واحدة كالآتي:

$$C_t = \beta_0 + \beta_1 Y_{dt} + \delta D_t + u_t \qquad t = 1, 2, \cdots, n$$

حيث ان:

C_t: الإنفاق الاستهلاكي الإجمالي في السنوات t.

Y_{dt}: مستوى الدخل المتاح في السنوات t.

D_t: متغير وهمي يوصف اثر الحرب والسلم كالآتي :

$$D_i = \begin{cases} 0 & \text{سنوات السلم} \\ 1 & \text{سنوات الحرب} \end{cases}$$ [*]

β_0: حد الكفاف

β_1: الميل الحدي للاستهلاك

δ: تعكس أهميه السلم وهي بذلك تعكس تأثير الحرب السلبي فى الإنفاق الاستهلاكي.

وبذلك فان متوسط الاستجابة :

الإنفاق الاستهلاكي في سنوات الحرب: $E(C_t / D = 1) = (\beta_0 + \delta) + \beta_1 Y_{dt}$

اما متوسط الاستجابة في فتره السلم: $E(C_t / D = 0) = \beta_0 + \beta_1 Y_{dt}$

(*) يمكن توصيف D كآلاتي:

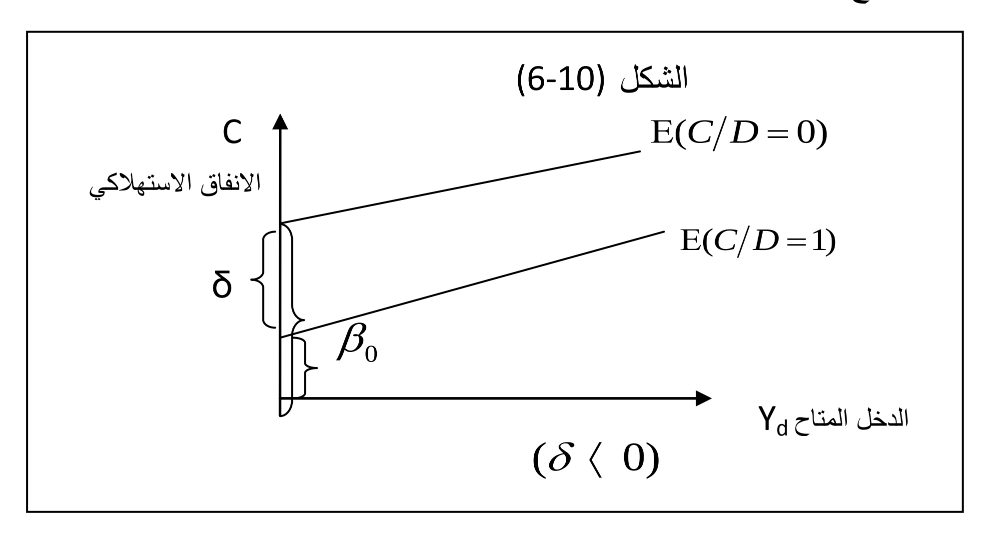
$$D_i = \begin{cases} 1 & \text{سنوات السلم} \\ 0 & \end{cases}$$

ويكمن الاختلاف في تفسير النتائج

وبذلك فان متوسط الاستجابةفي فتره السلم هي: $\delta\ E(C_t / D = 1) = \beta_0 + \delta + \beta_1 Y_{dt}$

وفي فتره الحرب: $E(C_t / D = 0) = \beta_0 + \beta_1 Y_{dt}$

وبذلك فعند مستويات الدخل المنخفضة، فان النفقات الاستهلاكية تنخفض بمقدار (δ) خلال سنوات الحرب. الشكل (10-6) يوضح ذلك:



اما اذا كان الافتراض ان ظروف وقت الحرب تقلل الميل الحدي للاستهلاك دون الحد الثابت في معادلة الاستهلاك وبذلك فان معادلة الانحدار لتمثيل دالة الاستهلاك في كلتا الفترتين :

$$C_t = \beta_0 + \beta_1 Y_{dt} + \gamma(Y_{dt} D_t) + u_t$$

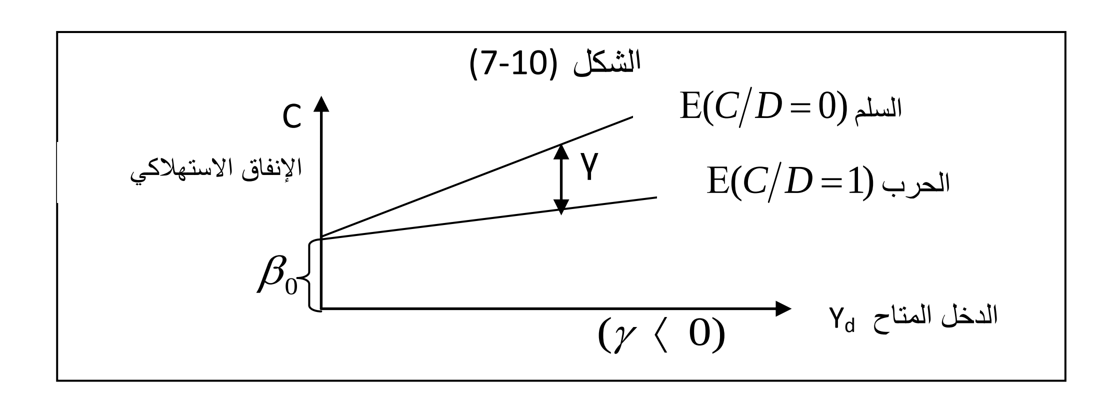
وبذلك فان متوسط الاستجابة للفترتين هو :

$$E(C_t / D = 1 \quad \text{(سنوات الحرب)} = \beta_0 + (\beta_1 + \gamma) Y_{dt}$$

$$E(C_t / D = 0 \quad \text{(سنوات السلم)} = \beta_0 + \beta_1 Y_{dt}$$

(ونتوقع بان γ سالبة).

والعلاقة المقدرة الناتجة يمكن تمثيلها بالمنحنيات الموجودة في الشكل (10-7) حيث ان دالة الاستهلاك في وقت الحرب ذات انحدار اقل، ولكن القاطع الرأسي نفسه كما هو الحال في وقت السلم.



(2-10)الانحدار في حالة أكثر من متغير نوعي واحد Regression in case of more than one dummyvariable

يمكن استخدام عدد كبير من المتغيرات النوعية (الصورية) بقدر ما يتطلب التحليل، شرط توافر عدد كاف من المشاهدات تسمح بتقدير المعادلة على سبيل المثال، في دراسة السلوك الاستهلاكي لأسر مختلفة فقد يعتمد مستوى الاستهلاك على عدد من السمات المميزة مثل وجود الأطفال من عدمهم، أو اذا كانت الاسرة تقطن في منزل ملك أو للإيجار، جنس رب الأسرة فيما اذا كان انثى أو ذكر وغيرها من المتغيرات النوعية الى جانب المتغير الكمي وهو الدخل المتاح فيمكن إضافة عدد من المتغيرات النوعية على وفق هدف الدراسة مع وجوب مراعاة الاتي:

- يجب ان تحدد عدد الفئات لكل متغير نوعي.
- تحدد الفئة المرجعية لكل متغير نوعي.

اما الخطوات الاخرى فهي نفسها في حالة متغير نوعي واحد.

وقد تنشأ حالات خاصة يمكن توضيحها من خلال دراسة التداخل بين المتغيرات الوهمية.

(3-10)التداخل بين المتغيرات الوهمية Interaction between dummies

في الفقرات السابقة تم توضيح ان اثر المتغيرات الوهمية يكون من خلال الاختلاف في المقطع العمودي (الصادي) (أو قد يكون من خلال الاختلاف في الميل في حالة وجود تداخل بين المتغير الوهمي والمتغير الكمي التوضيحي) كما انه يمكن استخدام عدد كبير من المتغيرات الوهمية بقدر ما يتطلب التحليل.

والسؤال هنا هل ان آثار المتغيرات الوهمية تكون مستقلة عن بعضها الآخر ؟

الإجابة عن هذا السؤال يكون من خلال المثال التالي:

مثال(4-10): دراسة الأجر الوظيفي Y دالة بدلالة الخبرة X ، الى جانب عوامل أخرى تتعلق بالإنتاجية والقابلية. لذا يضمن متغير الجنس (ذكراً أو أنثى) فضلا عن الحالة الزوجية (متزوج أو غير متزوج) وبذلك فان معادلة الانحدار:

$$Y = \beta_0 + \beta_1 X + \delta_1 D_1 + \delta_2 D_2 + u \qquad \ldots \qquad (9\text{-}10)$$

حيث ان:

$$D_1 = \begin{cases} 1 & \text{ذكر} \\ 0 & \text{انثى} \end{cases} \qquad \text{و} \qquad D_2 = \begin{cases} 1 & \text{غيرمتزوج} \\ 0 & \text{متزوج} \end{cases}$$

وبافتراض وجود تداخل بين المتغيرات الوهميه فان علاقة الانحدار تصبح:

$$Y = \beta_0 + \beta_1 X + \delta_1 D_1 + \delta_2 D_2 + \gamma (D_1 D_2) + u \qquad \ldots \qquad (10\text{-}10)$$

حيث ان :

δ_1 : تقيس اثر الجنس في حين δ_2 تقيس اثر الحالة الزوجية.

وγ : تقيس اثر كون الموظف ذكراً وغير متزوج.

وبذلك فان متوسط الاستجابة للأجر يتم توصيفها على وفق الآتي:

$$Y = \begin{cases} (\beta_0 + \delta_1 + \delta_2 + \gamma) + \beta_1 X & \text{ذكر غير متزوج} \\ (\beta_0 + \delta_2) + \beta_1 X & \text{انثى غير متزوجة} \\ \beta_0 + \delta_1 + \beta_1 X & \text{ذكر متزوج} \\ \beta_0 + \beta_1 X & \text{انثى متزوجة} \end{cases} \quad \cdots \quad (11-10)$$

(4-10) <u>الانحدار في حالة عدد فئات المتغيرالوهمي أكثر من ٢ : Regression in case of dummies with more than two categories</u>

هناك العديد من المتغيرات الوصفية تمتلك أكثر من صفتين. مثلاً الموقع الجغرافي أو مستوى التعليم أو فصول السنة، أو اشهر السنة. في هذه الحالة سيتم توليد متغير وهمي لكل فئة بشكل منفصل. هذا ويجب مراعاة عدم تضمين جميع المتغيرات الوهمية المتولدة في معادلة الانحدار منعاً لحدوث تعدد خطي تام وتسمى بمصيدة المتغير الوهمي(*). لذلك يجب توليد متغيرات وهمية بقدر (عدد الفئات – ١). ويكون المتغير الوهمي المحذوف ممثلاً لمجموعة المقارنة (Reference group)

ورياضياً لايهم أي فئة تعد الاساس للمقارنة. وسيتم عرض مجموعة امثلة لتوضيح ذلك.

مثال(5-10): دراسة الأجور (Y) دالة بدلالة عدد سنوات الخبرة (X) والتحصيل الدراسي. ونفترض ان العينة المدروسة تحصيلهم الدراسي هو : أقل من ثانوي، ثانوي، بكالوريوس، دبلوم عالي. وحيث ان عدد الفئات هو(4) لذا نولد المتغيرات الوهمية كالآتي :

$$D_3 = \begin{cases} 1 & \text{دبلوم} \\ 0 & \text{غيرذلك} \end{cases} \quad , \quad D_2 = \begin{cases} 1 & \text{جامعي} \\ 0 & \text{غيرذلك} \end{cases} \quad , \quad D_1 = \begin{cases} 1 & \text{ثانوي} \\ 0 & \text{غيرذلك} \end{cases}$$

وتكون فئة أقل من الثانوي هي الاساس للمقارنة

(*) يمكن التخلص من الوقوع في مصيدة المتغير الوهمي وذلك من خلال عدم تضمين مقطع صادي (ثابت) في معادلة الانحدار مع مراعاة تفسير المعلمات واختلافها عن حالة تضمين مقطع صادي).

وبذلك فان معادلة الانحدار للأجر حسب التحصيل والخبرة :

$$Y = \beta_0 + \beta_1 X + \delta_1 D_1 + \delta_2 D_2 + \delta_3 D_3 + u$$

وان متوسط الاستجابة للأجر:

$$\mathrm{E}(Y) = \begin{cases} \beta_0 + \beta_1 X & \text{اقل من ثانوي} \\ (\beta_0 + \delta_1) + \beta_1 X & \text{ثانــوي} \\ \beta_0 + \delta_2 + \beta_1 X & \text{بكالوريوس} \\ (\beta_0 + \delta_3) + \beta_1 X & \text{دبلوم عالي} \end{cases}$$

حيث ان:

δ_3: يقيس الفرق في متوسط الأجر لموظف حاصل على دبلوم عالي مقارنة بشخص آخر تحصيله أقل من ثانوي.

δ_2: يقيس الفرق في متوسط الأجر لموظف حاصل على بكالوريوس مقارنة بأقل من ثانوي

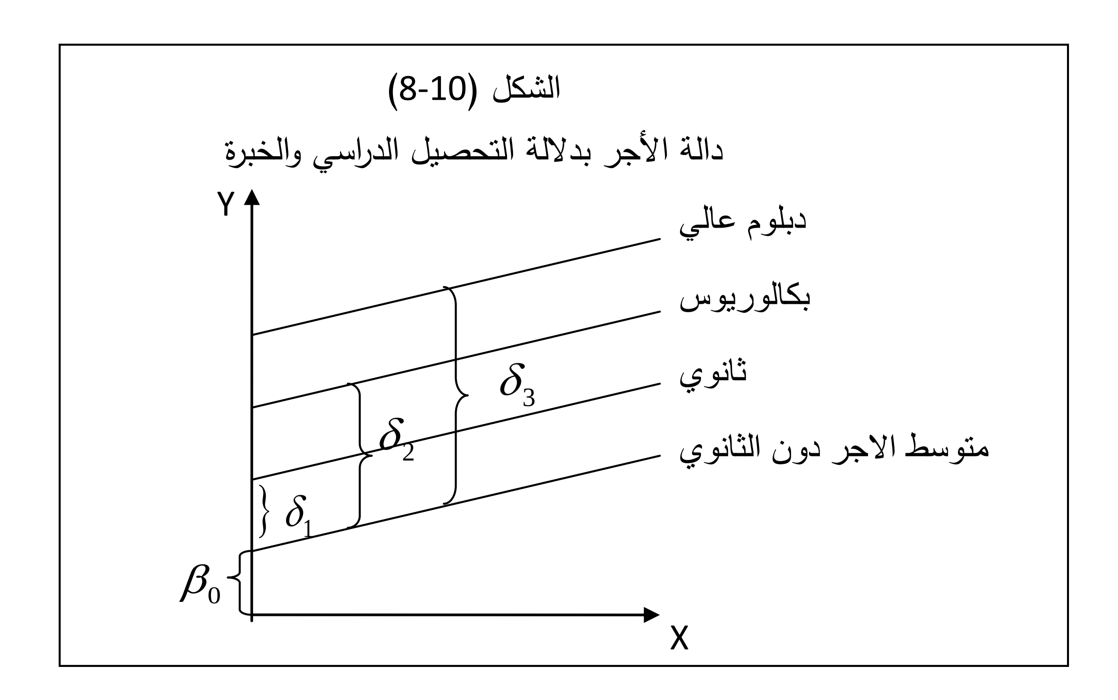
δ_1: يقيس الفرق في متوسط الأجر لموظف حاصل على ثانوية.

β_1: يمثل الأجر الأساسي لموظف بدون خبرة وتحصيله اقل من ثانوي.

ويمكن توضيح ذلك من خلال الشكل (8-10)

الشكل (8-10)

دالة الأجر بدلالة التحصيل الدراسي والخبرة



وهناك استخدامات إضافية للمتغيرات الوهمية أيضاً ويمكن توضيحها بالأمثلة:

مثـال(6-10): فـي دراسـة الموسـمية : أن التقلبـات الموسـمية مـن بـين العوامـل التـي تـؤثر فـي الظـواهر الاقتصـادية. علـى سـبيل المثـال فـي دراسـة الطلـب علـى الملابـس الصـوفية أو الملابـس الثقيلـة بوجـه عـام يزداد عند دخول الموجات الباردة في فصل الشتاء في حين يقل الطلب عليها في فصل الصيف، كما ان الطلب على المشروبات الباردة والايس كريم يزداد في الموجات الحارة في فصل الصيف ويقل الطلب عليها في الموجات الباردة من فصل الشتاء. في حين الطلب على الحلوى والهدايا يزداد في موسم الاعياد بصورة ملحوظة بالمقارنة مع الأوقات الاخرى من السنة .

ولدراسـة اثـر الموجـات الحـارة فـى الطلـب علـى المشـروبات البـاردة يـتم إدخـال متغيـر وهمـي D يـتم توصيفه على وفق الآتي :

$$D = \begin{cases} 1 & \text{الصيف} \\ 0 & \text{غيرذلك} \end{cases}$$

وبذلك تكون معادلة الطلب على المثلجات على وفق الآتي : $Y = \beta_0 + \beta_1 X + \delta D + u$

وبذلك فان المعلمةδ تقيس الاختلاف في الطلب في موسم الموجة الحارة في الصيف عن باقي ايام السنة

وبذلك فان التأثير يظهر من خلال المقطع الثابت كما يمكن ان يكون التأثير من خلال الميل الحدي :

$$Y_i = \beta_0 + \beta_1 X_i + \gamma (D_i X_i) + u_i$$

أو قد يكون التأثير من خلال المقطع الثابت والميل الحدي : $Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \gamma X_i + v_i$

ولإظهار اثر المواسم الأربعة (صيف ، خريف ، شتاء ، ربيع) فيتم من خلال إدخال عدد من المتغيرات الوهمية D_1 ،D_2 ،D_3 ،D_4 لنموذج لا يحوي على مقطع صادي ويكون توصيف المتغيرات الوهمية على وفق الآتي :

$$D_4 = \begin{cases} 1 & \text{صيف} \\ 0 & \text{غيرذلك} \end{cases} , \quad D_3 = \begin{cases} 1 & \text{خريف} \\ 0 & \text{غيرذلك} \end{cases} , \quad D_2 = \begin{cases} 1 & \text{شتاء} \\ 0 & \text{غيرذلك} \end{cases} , \quad D_1 = \begin{cases} 1 & \text{ربيع} \\ 0 & \text{غيرذلك} \end{cases}$$

وتكون معادلة الانحدار: $Y = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + u$

أو يكـون النمـوذج بإدخـال ثلاثـة متغيرات وهميـة وحـذف المتغيـر الـوهمي الـذي يعد موسم المقارنـة مـع المقطع الصادي، ويكون موسم المقارنة خيارياً يعتمد على الباحث. فاذا جعلنا فصل الصيف هو الفصل الذي تتم على وفقه المقارنة يحذف D_1.

وتكون معادلة الانحدار : $Y_i = \beta_0 + \delta_2 D_{2i} + \delta_3 D_{3i} + \delta_4 D_{4i} + u_i$

اما اذا رغب الباحث في جعل فصل الخريف هو فصل المقارنة فتكون معادلة الانحدار :

$$Y = \beta_0 + \delta_1 D_{i1} + \delta_3 D_{3i} + \delta_4 D_{4i} + u_i$$

وهكذا . . .

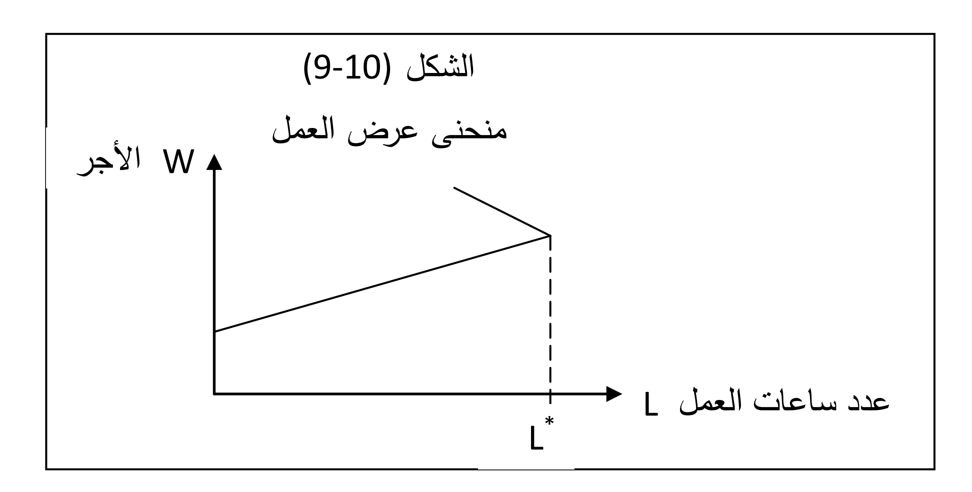
ومن ضمن الاستخدامات المهمة الاخرى للمتغيرات الوهمية يكون بدراسة المنحنى المنكسر.

مثال(7-10): يستخدم المتغير الوهمي لدراسة المنحنى المنكسر. وذلك من خلال قياس التغير في الميل بعد مستوى معين.

لدراسة منحنى عرض العمل، فارتفاع الأجر يؤدي الى زيادة عدد ساعات العمل الى حد معين. وبعد هذا المستوى يؤدي ارتفاع الأجر الى تقليل ساعات العمل، وكما في الشكل (10-9).

اذ ان معادلة الانحدار التي تصف عرض العمل هي:

$$W = \beta_0 + \beta_1 D + u \quad ... \quad (13\text{-}10)$$



الشكل (10-9)
منحنى عرض العمل

$$W = \beta_0 + \beta_1 D + u$$

$$D = \begin{cases} 1 & (L > L^*)\ L^* \text{ عدد الساعات أكبر من} \\ 0 & \text{غير ذلك} \end{cases}$$

مثال(8-10): كما ويعد المتغير الوهمي مؤشراً مهما لدراسة المتغيرات الرقمية. في الحالات التي لا تتاح بيانات كافية عن هذه المتغيرات. مثلاً دراسة دالة الادخار لعدد من الإفراد، فان العمر يعد احد المتغيرات التوضيحية تؤثر فى الادخار اذ يعتقد بانه كلما تقدم الفرد في السن زادت حكمته وبالتالي تزداد مقدرته على الادخار. ولكن لا يكون الاختلاف في العمر بعام أو عامين فقد يكون ذلك ليس بالتأثير الكبير الذي يؤثر فى اتخاذ قرارات الادخار وانما الاختلافات الكبيرة هي التي تؤثر. فاذا كانت أعمار العينة تترأوح بين 15 سنة و 65 سنة فقد يكون من المناسب تقسيمهم الى ثلاث مجاميع:

المجموعة الأولى: تضم الإفراد الذين تترأوح أعمارهم بين15-30 سنة

المجموعة الثانية: تضم الإفراد الذين تترأوح أعمارهم بين 31-45 سنة

المجموعة الثالثة: تضم الإفراد الذين تترأوح أعمارهم بين 46-65 سنة

$$S_t = \beta_0 + \beta_1 Y_t + \delta_1 D_{1t} + \delta_2 D_{2t} + \gamma_1 D_{1t} Y_t + \gamma_2 D_{2t} Y_t + u_t \quad . . . \quad (14-10)$$

وبذلك فان دالة الادخار والتي يعتقد بأنها تتأثر بمستوى الدخل وعمر الفرد يمكن صياغتها على وفق الآتي:

S: الادخار ، Y: الدخل

$$D_1 = \begin{cases} 1 & \text{الفرد من المجموعة الثانية} \\ 0 & \text{غير ذلك} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{الفرد من المجموعة الثالثة} \\ 0 & \text{غير ذلك} \end{cases}$$

وبذلك تكون فئة الأساس هي الفئة الأولى .

كما ويمكن استخدام المتغيرات الوهمية لدراسة قطاعات مختلفة عبر عدد من السنوات وهناك استخدامات أخرى واسعة ومتعددة، تعد خارج نطاق منهج الكتاب.

(5-10) اختيار المعنوية في نماذج تحتوي متغيرات وهمية:

" Testing hypolhesis of effects of Dummy variables"

(1-5-10) اختيار معنوية آثار متغير وهمي منفرد(اختبار معنوية)

" Testing hypothesis of intercept Dummy "

مع افتراض ان المتغير العشوائي يتوزع توزيعاً طبيعياً أو ان حجم العينة كبير فان اختبار ستيودنت t يمكن إتباعه لاختبار معنوية اثر احد المتغيرات الوهمية:

$$\mathrm{H}_0 : \delta = 0$$

$$vs. \quad \mathrm{H}_1 : \delta \neq 0 \quad or \quad \mathrm{H}_2 : \delta \rangle\ 0$$

يمكن استخدام t :

$$t^* = \frac{\hat{\delta}}{s.e(\hat{\delta})}$$

ومقارنتها بالقيم الجدولية لطرفين في الحالة الأولى ولطرف واحد للحالة الثانية.

وبذلك فان الفشل في معنوية (δ) يكون عدم وجود اختلاف معنوي للصفة التي يمتلكها المتغير الوهمي . وبالعودة الى المثال (10-1) لاختبار فيما إذا كان الموقع ذا دلالة معنوية على سعر السكن فيكون ذلك من خلال اختبار معنوية (δ) اذ تكون فرضية العدم $\mathrm{H}_0 : \delta = 0$ ضد $\mathrm{H}_1 : \delta \neq 0$ وبذلك فان الفشل في معنوية (δ) يمكن تفسيره بان الموقع متكامل الخدمات ليس بذي أهمية لاختلاف السعر للسكن. ويتم ذلك باستخدام المختبر (t).

(2-5-10) اختيار المعنوية المشتركة لآثار عدد من المتغيرات الوهمية .

" Testing for the joint signficancy"

كما ويمكن اختبار الأهمية المشتركة للمتغيرات الوهمية، بافتراض وجود متغيرين وهميين.

معادلة الانحدار هي: $Y = \beta_1 + \beta_2 X + \delta_1 D_1 + \delta_2 D_2 + \gamma \; D_1 D_2 + u$

ولاختبار المعنوية المشتركة فان فرضية العدم :

$$\mathrm{H}_0 : \delta_1 = \delta_2 = \gamma = 0$$

والبديلة : $vs. \quad \mathrm{H}_1 : \delta_1 \neq \delta_2 \neq \gamma \neq 0$

ويتم استخدام المختبر F على وفق القانون:

$$F^* = \frac{(e'e_r - e'e_{Ur})/J}{e'e_{Ur}/(n-k)}$$

J: عدد المتغيرات الوهمية = عدد الفرضيات المشتركة.

k: عدد المعلمات المطلوب تقديرها في نموذج الانحدار غير المقيد.

ويتم الحصول على مجموع مربعات الأخطاء للنموذج المقيد $(e'e_r)$من تقدير النموذج :

$$Y = \beta_1 + \beta_2 X + error$$

اما مجموع مربعات الأخطاء للنموذج غير المقيد ($e'e_{Ur}$) من تقدير النموذج:

$$Y = \beta_1 + \beta_2 X + \delta_1 D_1 + \delta_2 D_2 + \gamma \; D_1 D_2 + e$$

ولاختبار أهمية الموسمية في البيانات لابد من اختبار المعنوية المشتركة لجميع المتغيرات الوهمية التي تمثل الموسمية.

ففي النموذج: $Y = \beta_0 + \delta_2 D_2 + \delta_3 D_3 + \delta_4 D_4 + u$

حيث D_2 ، D_3 ، D_4 متغيرات وهمية تمثل الموسمية وباعتبار الموسم الأول هو موسم المقارنة.

ولاختبار الموسمية لابد من اختبار الفرضية:

$$\mathrm{H}_0 : \delta_2 = \delta_3 = \delta_4 = 0$$

H_0: عدم وجود الموسمية

والتي تستوجب استخدام المختبر F عوضاً عن t ، حيث ان النموذج غير المقيد هو :

$$Y = \beta_0 + \delta_2 D_2 + \delta_3 D_3 + \delta_4 D_4 + u$$

في حين يكون النموذج المقيد:

$$Y = \beta_0 + u$$

حيث ان الموسمية هي مركبة مشتركة لجميع المتغيرات الوهمية التي تمثل الموسمية.

$$F^* = \frac{(e'e_r - e'e_{Ur})/3}{e'e_{Ur}/(n-4)}$$

وبالمنهجية نفسها فان اختبار البيانات الشهرية تكون باستخدام F اذ نختبر المعنوية المشتركة لجميع المتغيرات الوهمية التي تمثل البيانات الشهرية.

(3-5-10) اختبار تساوي معادلتي انحدار باستخدام المتغيرات الوهمية.

Testing for the equality of two regression using Dummy variable.

يمكن استخدام المتغيرات الوهمية لدراسة تسأوي الآثار لمعادلة النموذج ككل.

ففي دراسة قيمة السكن كدالة بدلالة (مساحة السكن) وموقعه فان متوسط الاستجابة تحدد على وفق:

$$E(P_t) = \begin{cases} (\beta_1 + \delta) + (\beta_2 + \gamma)S_t \equiv \alpha_1 + \alpha_2 S_t & \text{الموقع متكامل الخدمات} \\ \beta_1 + \beta_2 S_t & \end{cases}$$

في هذه الحالة فقد تم السماح للمقطع الثابت والميل بالاختلاف $\alpha_2 \neq \beta_2 \& \alpha_1 \neq \beta_1$ وبذلك فان الانحدار في كلتا الحالتين مختلف تماماً وكما موضح في الشكل(5-10)

وعليه يكون تقدير المعلمات في النموذجين بشكل منفصل (α_2 ، α_1) في النموذج في الموقع المرغوب و β_2 ، β_1 في الموقع الثاني.

ولاختبار هل توجد فروق معنوية بين الانحدار في الموقعين (متكامل الخدمات ، موقع آخر) يمكن دمج البيانات في عينة واحدة بتضمين المتغير الوهمي وتقدير النموذج

$$P = \beta_0 + \beta_1 S + \delta D + \gamma (SD) + u$$

باستخدام العينة المدمجة واختبار الفرضية:

$H_0 : \delta = \gamma = 0$ والتي تدل على عدم وجود اختلاف في سعر السكن بغض النظر عن الموقع

vs

$H_1 : \delta \neq \gamma \neq 0$ والتي تدل على وجود اختلاف في سعر السكن اعتماداً على [illegible] الموقع

ونلاحظ في حالة تحقق فرضية العدم فان النموذج المستخدم:

$$P = \alpha_1 + \alpha_2 S_t + e_t \quad \text{الموقع الكامل}$$

$$P = \beta_1 + \beta_2 S_t + e_t \quad \text{الموقع الآخر}$$

وهو النموذج المقيد.

حيث ان :

$$\alpha_1 = \beta_1 \Leftarrow \ \delta = 0$$

و $$\alpha_2 = \beta_2 \Leftarrow \ \gamma = 0$$

في حين النموذج غير المقيد هو: $P_t = (\beta_1 + \delta) + (\beta_2 + \gamma) S_t + e_t$

ويتم اختبار الفرضية على وفق اختبار F كما في الفقرة السابقة. وهو ما يعرف (chow test).

مثال (9-10): افترض بيانات عن الاستثمار الإجمالي (Y) في عدد الآلات لشركتين مثلاً جنرال الكترك (I) ويستتج هأوس (II) للسنوات (1935-1954) بأسعار 1947 كدالة بدلالة قيمة الشركتين للمدة نفسها X_1 ورصيد رأس المال في الشركتين للمدة نفسها X_2 ، وبذلك فان دالة الانحدار هي:

$$\begin{cases} Y_{I} = \alpha_0 + \alpha_1 X_{1I} + \alpha_2 X_{2I} + e_{I} & n_{I} = 20 \\ Y_{\Pi} = \gamma_0 + \gamma_1 X_{1\Pi} + \gamma_2 X_{2\Pi} + e_{\Pi} & n_{\Pi} = 20 \end{cases}$$

وعند دمج الشركتين بعينة واحده فيكون حجم العينة = 40

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

وبذلك يكون دمج العينة هو الدالة المقيدة (افتراض تسأوي الميلين والمقطعين الثابتين) ونحسب مجموع مربعات البواقي للدمج $e'e_r$.

في حين مجموع مربعات البواقي للانحدار غير المقيد هو $(e'e)_r = (e'_{\Pi}e_{\Pi} + e'_{I}e_{I})$.

ثم نستخدم اختبار F :

$$F^* = \frac{(e'e_r - e'e_{Ur})/3}{e'e_{Ur}/(n_{I} + n_{\Pi} - 2k)}$$

ويمكن أيضاً استخدام (chow - test) من خلال إدخال متغير وهمي D يوصف كآلاتي :

$$D = \begin{cases} 1 & \text{لبيانات الشركة وستنج هأوس} \\ 0 & \text{للشركة الأخرى} \end{cases}$$

وبذلك تكون معادله الانحدار :

$$Y = \beta_0 + \delta_1 D + \beta_2 X_1 + \delta_2 (DX_1) + \beta_3 X_2 + \delta_3 (DX_2) + e_t$$

وهي تمتلك معادله الانحدار غير المقيدة، نحسب مجموع المربعات لبواقي العلاقة: $(e'e_{Ur})$

كما يتم حساب مجموع مربعات البواقي للعلاقة المقيدة: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$

ثم يتم استخدام المختبر F .

وبافتراض ان تقدير العلاقة للبيانات المدمجة (40) مشاهدة اعطت المعلومات التالية:

$$e'e_r = 16563.00 \quad , \quad \begin{cases} \hat{Y} = 17.872 + 0.0152X_1 + 0.1436X_2 \\ t: \quad (2.544) \quad (2.452) \quad (7.719) \end{cases}$$

وكذلك معادلة التقدير باستخدام المتغيرات الوهمية هي:

$$\hat{Y} = -9.9563 + 9.4469D + 0.0266X_1 + 0.0263(DX_1) + 0.1517X_2 - 0.0593(DX_2)$$
$$t: \quad (0.421) \quad (0.328) \quad (2.265) \quad (0.767) \quad (7.837) \quad (-0.507)$$

$$e'e_{Ur} = 14989.82$$

وبتطبيق F:

$$F^* = \frac{(e'e_r - e'e_{Ur})/J}{e'e_{Ur}/(n-k)} = \frac{(16563.00 - 14989.82)/3}{14989.82/(40-6)} = 1.1894$$

$$F_{c(3,34,0.95)} = 2.8826$$

وبذلك فان القرار يكون:

لا ترفض فرضية العدم. وبذلك فان دالة الاستثمار في الشركتين متسأوية (غير مختلفة).

أي لا اختلاف في استثمار الشركتين.

(6-10)تفسير المتغيرات الوهمية في معادلة الانحدار نصف اللوغاريتمية Interpretation of the dummy variable in semi log regression

في حاله النماذج نصف اللوغاريتمية (log – lin):

$$LnY = \beta_0 + \beta_1 X$$

عندما تتضمن العلاقة نصف اللوغاريتمية متغيراً وهمياً:

$$LnY = \beta_0 + \beta_1 X + \delta_1 D$$

مثلاً Y : الدخل للتدريسي

X : سنوات الخدمة التدريسية.

D : شهادة الدكتوراه = 1 وماجستير = 0

δ_1 : التغير النسبي في Y لتغير مطلق في D

عند زيادة سنوات الخدمة سنة واحدة فان متوسط الدخل يمثل (β_1)

اما فيما يخص تفسير(δ_1) فيكون بأخذ العدد المقابل بأساس e للمعلمة $\delta_1\delta$ وطرحها من الواحد.

$$\left(1 - e^{\delta_1}\right) \equiv$$

بعبارة اخرى فان التغيير النسبي في متوسط Y نسبة الى الشهادة يتمثل بـ $\left\{ 1 - (anti \ln(\hat{\delta})) \right\}$

مثـال(10-10): افترض معادلـة التقدير لـدخل التدريسـي كدالـة بسـنوات الخدمـة (X) وان العينـة مجموعـة تدريسيين بعضـهم حاصل على شهادة دكتوراه والآخرين شهادة ماجستير.

$$Ln\hat{Y} = 2.9298 + 0.0546X_1 + 0.1341D$$

$$t: \quad (481.524) \quad (48.3356) \quad (27.225)$$

$R^2 = 0.9958$ $\quad$ d =2.51

يتوضح من النتائج ان معلمة الانحدار هي (0.0546) بمعنى ان مقدار الزيادة السنوية لـ (لوغاريتم الدخل) هي % 5.46.

أي ان متوسط الدخل يزداد بمقدار:

$$\begin{Bmatrix} 1-(anti\ln(0.1341)) = (1-1.1435) \\ = 0.1435 \end{Bmatrix}$$

أي بمقدار %14.35 مع ثبات سنوات الخبرة للتدريسي الحاصل على شهادة دكتوراه.

(7-10)المتغيرات الوهمية ومشكلة عدم تجانس الأخطاء Dummy variables and hetroscedastisity

عند دمج مشاهدات(فترتين أو أكثر) باستخدام المتغيرات الوهمية يجب افتراض ان التباين في كـل فتـرة هـو نفسـه فـي الفتـرة الأخـرى $\{ \operatorname{var}(u_i) = \sigma^2 \quad \forall \quad i = 1,2,\cdots \}$ أي افتـراض تجـانس التباين للأخطاء.

اما في حالة عدم تحقق فرضية التجانس، فان نتائج التقدير قد تظهر الآتي:

(ان المقطع والميل يكون غير مهم إحصائياً بالرغم من ان معلمة المتغير الوهمي تكون معنوية).

وعليه فان لاختيار التغيرات الهيكلية باستخدام المتغيرات الوهمية (اختبار chow) لابد من التحقق بعدم وجود مشكلة عدم التجانس.

(8-10) المتغيرات الوهمية ومشكلة الارتباط الذاتي Dummy variables and autocorrelation

عند استخدام دراسات السلاسل الزمنية لمدتين (n_1) و (n_2) فيتم استخدام المتغير الوهمي:

$$D = \begin{cases} 0 & \text{للمدة الأولى} \\ 1 & \text{للمدة الثانية} \end{cases}$$

وبذلك فان معادلة الانحدار:

$$Y = \beta_0 + \beta_1 X + \delta D + \lambda(XD) + u$$

بافتراض ان المتغير العشوائي يمتلك AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_t$$

ε_t: متغير عشوائي ضوضاء بيضاء

$$\begin{cases} \mathrm{E}(\varepsilon_t) = 0 \\ \mathrm{var}(\varepsilon_t) = \sigma^2 \\ \mathrm{cov}(\varepsilon_t \varepsilon_{t-s}) = 0 \end{cases}$$

بذلك لابد من تحويل المشاهدات بالشكل الذي يخلص الأخطاء من مشكلة الارتباط الذاتي وان وجود المتغير الوهمي يولد مشكلة تحويل المشاهدات بالنسبة للمتغير الوهمي لأن قيمها اما (1) أو (صفر). يتم التحويل على وفق الآتي:

حيث ان:

$$D = \begin{cases} 0 & \text{للمدة الأولى} \\ 1 & \text{للمدة الثانية} \end{cases}$$

يتم تحويلها :

$$D^* = \begin{cases} 0 & \text{المدة الأولى} \\ \begin{cases} \dfrac{1}{1-\rho} & \text{المشاهدة الأولى} \\ 1 & \text{للمشاهدة الأخرى} \end{cases} \end{cases} \quad \cdots \quad (15-10)$$

أي:

$$D^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & \dfrac{1}{1-\rho} & 1 & \cdots & 1 \end{bmatrix}$$

$$n_1 \qquad n_1 + 1 \qquad n_1 + n_2$$

$$(D_t X_t)^* = \begin{cases} 0 & \text{المدة الأولى} \\ \text{المدة الثانية} \begin{cases} X_t & \text{المشاهدة الأولى} \\ D_t X_t - D_t X_{t-1} = X_t - \rho X_{t-1} & \text{للمشاهدات الأخرى} \end{cases} \end{cases} \quad \cdots \quad (16-10)$$

مثال (11-10): انحدار Y (الدخل الشخصي) دالة بدلالة X (سنوات الخدمة) ، D_1، D_2 الحالة التعليمية، ومعادلة التقدير:

$$\hat{Y} = 1360.638 + 375.5X + 1592.1D_1 + 4707.4D_2 + 515.3D_1X + 339.5D_2X$$

م/1) حدد دالة الاستجابة لكل مستوى من مستويات الدالة التعليمية، وفسر مدلول العلاقات المقدرة.

2) هل ان خطوط الانحدار متطابقة ؟

3) هل يوجد تداخل بين المتغيرات D_1 ، D_2 والمتغير التوضيحي X. بافتراض توافر المعلومات التالية:

s.o.v	d.f	SS
$ESS(X,D_1D_2,D_1X,D_2X)$	5	261.95
$ESS(X)$	1	95.605
$ESS(X,D_1,D_2)$	3	246.87
$ESS(D_1D_2,D_1X,D_2X / X)$	4	166.345
$ESS(D_1X,D_2X / XD_1,D_2)$	2	15.080
Error	14	185.33
Total	19	447.28

الحل:

(1) دالة الاستجابة هي: $\mathrm{E}(Y) = \hat{\beta}_0 + \hat{\beta}_1 X + \delta_1 D_1 + \delta_2 D_2 + \lambda_1 (D_1 X) + \lambda_2 (D_2 X)$

وبذلك فان دالة الاستجابة للمستويات المختلفة:

أساس المقارنة:

$$\mathrm{E}\left(Y \middle/ \begin{matrix} D_1 = 0 \\ D_2 = 0 \end{matrix} \right) = \hat{\beta}_0 + \hat{\beta}_1 X = 1360 + 375X$$

المستوى الثاني

$$\mathrm{E}(Y/D_2 = 1, \quad D_1 = 0) = (1360 + 4707.4) + (375 + 339.5)X$$

المستوى الثالث

$$\mathrm{E}(Y/D_1 = 1, \quad D_2 = 0) = (1360 + 1592) + (375 + 515.3)X$$

(2) لاختبار تطابق خطوط الاستجابة

$$\mathrm{H}_0 : \delta_1 = \delta_2 = \lambda_1 = \lambda_2 = 0$$

$$vs. \quad \mathrm{H}_1 : \delta_1 \neq \delta_2 \neq \lambda_1 \neq \lambda_2 \neq 0$$

$$F^* = \frac{ESS(D_1D_2, D_1X, D_2X / X)/4}{RSS(X, D_1, D_2, D_1X, D_2X)/(n-6)} = \frac{41.58625}{13.23785714} = 3.14$$

$$F_{c(4,14,0.95)} = 3.11$$

خطوط الاستجابة غير متطابقة ⟵ المتغيرات الوهمية معنوية.

(3) لاختبار التداخل بين المتغيرات الوهمية والمتغير الكمي X ، تصاغ فرضية العدم:

$$H_0 : \lambda_1 = \lambda_2 = 0$$

$$vs. \quad H_1 : \lambda_1 \neq \lambda_2 \neq 0$$

$$F^* = \frac{ESS(D_1X, D_2X / X, D_1, D_2)/2}{RSS(X, D_1, D_2, D_1X, D_2X)/(n-6)} = \frac{15.08/2}{185.33/14} = \frac{7.54}{13.23} = 0.569$$

$$F_{c(2,14,0.95)} = 3.74$$

وبذلك فان القرار هو لا أهمية للتداخل.

مثال(10-12): دراسة الادخار في فترتين (اختلاف سياستين) مثلاً

تطبيق السياسة I $\quad Y_i = \lambda_1 + \lambda_2 X_i + u_i \qquad i = 1,2,\cdots,n_1$

$$Y_i = \gamma_1 + \gamma_2 X_i + v_i \qquad i = 1,2,\cdots,n_2$$

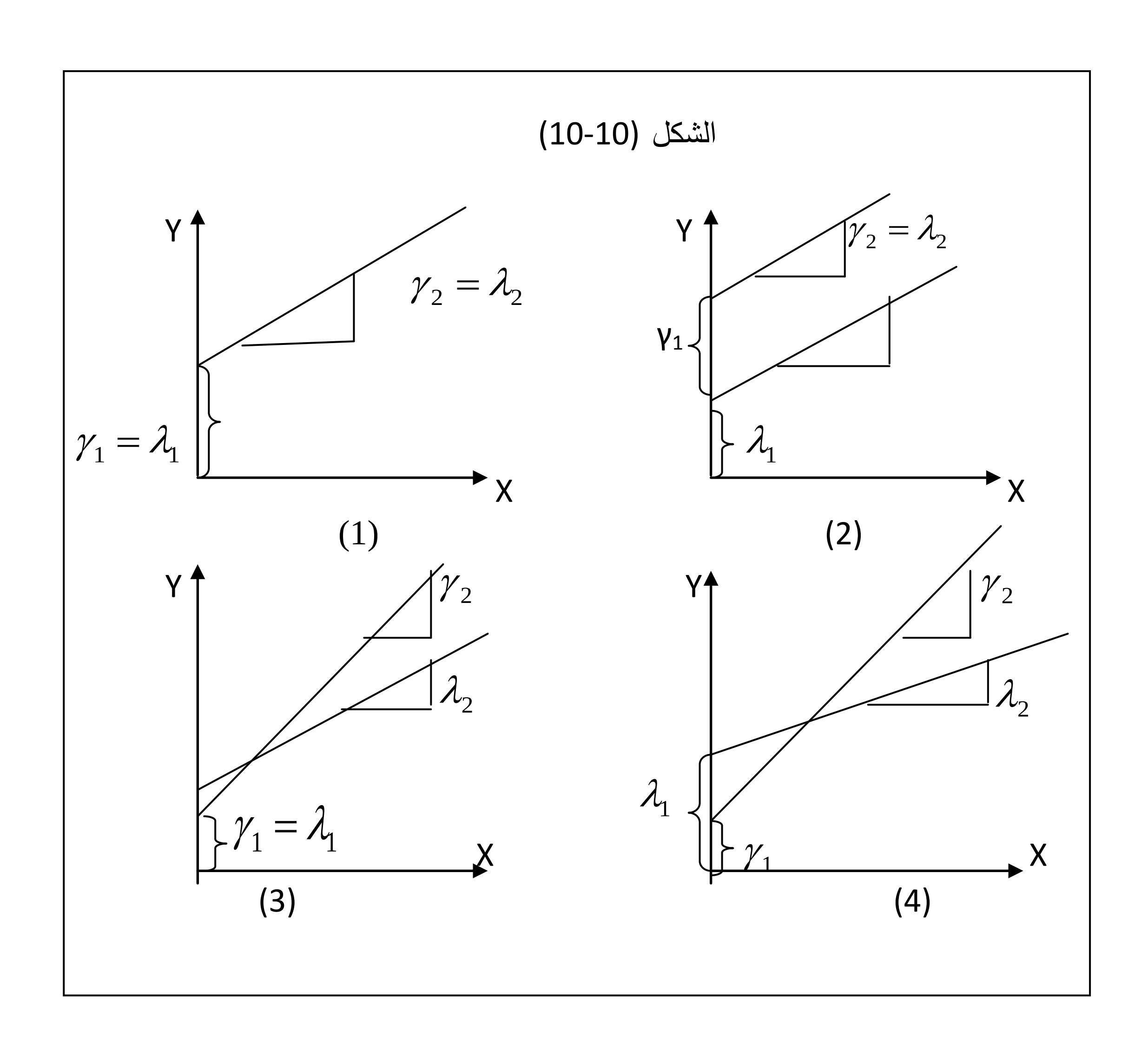
n_2، n_1 ليست بالضرورة متسأوية.

تنتج الحالات التالية:

1. معادلتا الانحدار متطابقة: $\lambda_1 = \gamma_1$ و $\lambda_2 = \gamma_2$ الانحدار متماثلان
2. الانحدار متوازيان: $\lambda_1 \neq \gamma_1$ و $\lambda_2 = \gamma_2$ الاختلاف في المقطع الصادي
3. Concurrent reg $\lambda_1 = \gamma_1$ و $\lambda_2 \neq \gamma_2$ الاختلاف في الميل
4. الانحدار مختلفان تماماً $\lambda_1 \neq \gamma_1$ و $\lambda_2 \neq \gamma_2$ في الميل والمقطع الصادي.

ولاختبار التغيرات الهيكلية يستخدم (اختبار chow) وكبديل لاختبار chow يتم استخدام المتغيرات الوهمية بعد تحديد نقطة التغير.

ويمكن توضيح الاحتمالات بالشكل البياني(10-10):


الشكل (10-10)
Y
X
$\gamma_2 = \lambda_2$
$\gamma_1 = \lambda_1$
(1)
Y
X
$\gamma_2 = \lambda_2$
γ_1
λ_1
(2)
Y
X
γ_2
λ_2
$\gamma_1 = \lambda_1$
(3)
Y
X
γ_2
λ_2
λ_1
γ_1
(4)

مثال(10-13): لدراسة العلاقة بين مقدار الراتب السنوي (Y) ، وعدد سنوات الخدمة (X_1) والخبرة الادارية (D) لعشرين موظف في مؤسسة معينة.

Y	D (متغير نوعي)		X_1 (متغير كمي)	رقم الموظف
3250	ليست له خبرة	0	5	1
1600	ليست له خبرة	٠	1	2
7500	ليست له خبرة	٠	16	3
13750	له خبرة	١	10	4
6000	له خبرة	1	5	5
8700	له خبرة	1	1	6
11350	له خبرة	1	3	7
3000	ليست له خبرة	0	4	8
15700	له خبرة	1	15	9
113500	له خبرة	1	6	10
١٢٢٥٠	له خبرة	1	8	11
700	ليست له خبرة	0	2	12
10250	له خبرة	1	4	13
3500	ليست له خبرة	0	8	14
4500	ليست له خبرة	0	10	15
16350	له خبرة	1	13	16
3800	ليست له خبرة	0	6	17
9800	له خبرة	1	3	18
10800	له خبرة	1	2	19
23300	ليست له خبرة	0	3	20

تمثيل المتغيرات النوعية باستخدام المتغيرات الوهمية.

المتغير النوعي D ذو فئتين:

$$D = \begin{cases} 1 & \text{له خبرة} \\ 0 & \text{ليست له خبرة} \end{cases}$$

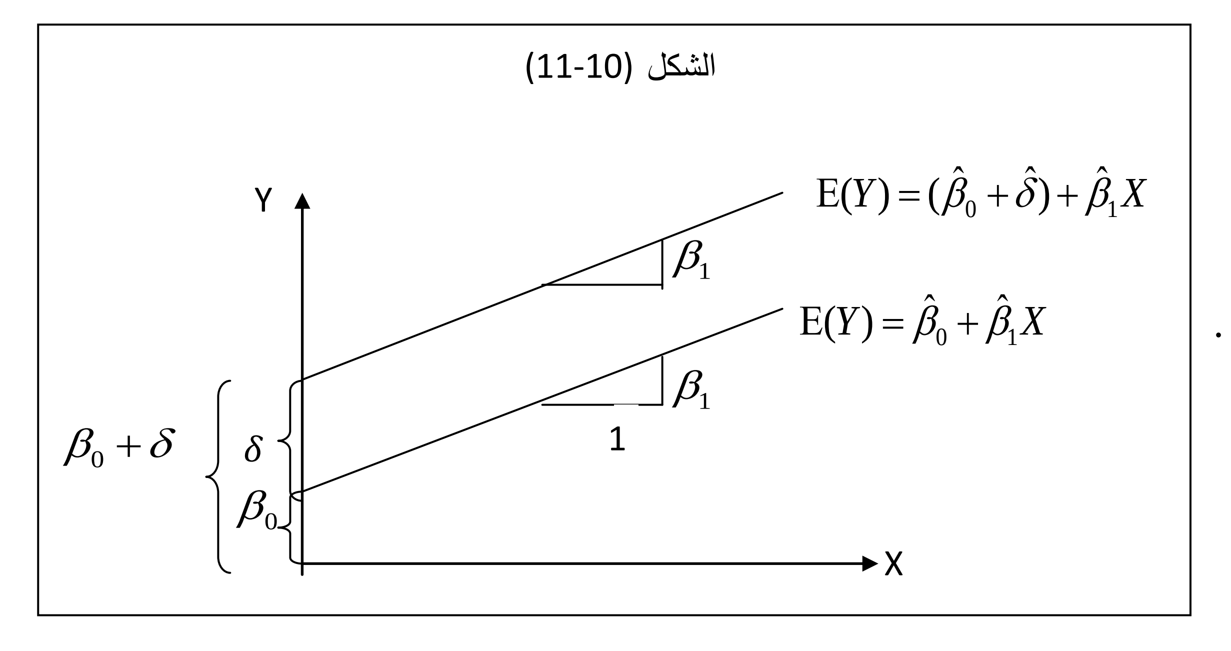
دالة الاستجابة: $E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\delta}\, D$

دالة الاستجابة لموظف له خبرة إدارية:

$$E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\delta}$$
$$= (\hat{\beta}_0 + \hat{\delta}) + \hat{\beta}_1 X$$

دالة الاستجابة لموظف ليست له خبرة إدارية: $E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X$

الشكل (11-10)



δ : هو مقياس للفرق في التأثير في نوع الخبرة الإدارية

دالة الاستجابة هي دالة خطية لعدد سنوات الخدمة.

بافتراض عدم وجود تداخل بين المتغير الوهمي D والمتغير الكمي X ، وباستخدام طريقة المربعات الصغرى اعطت النتائج التالية:

$$\hat{Y} = 4423.144 + 206.213X + 15032.775D$$

$t:$	(0.168)	(1.385)
$s.e:$	1226.658	10856.675

جدول تحليل التباين:

s.o.v	SS	d.f	MS	F
ESS(X,D)	1142860793.148	2	571430396.574	0.980
Error	9910466206.852	17	582968600.403	
Total	11053327000.000	19		

($F_c = 4.45$) لمستوى احتمال 1%.

نرى ان نتائج جدول التباين تشير الى عدم معنوية المعلمات مجتمعة. وكذلك تشير نتائج t الى عدم معنوية كل معلمة منفردة.

ان معنوية (δ) تدل على ان خط الانحدار للموظف الذي له خبرة والاخر الذي ليست له خبرة غير متطابقين. وان زيادة عدد سنوات الخدمة سنة واحدة ستزيد الراتب السنوي بمقدار 206.213 دينار(مع ثبات D).

$\hat{\delta} = 15032.775$: تعني التغير في راتب الموظف اعتماداً على خبرته.

كما ان مجال الثقة باحتمال 99% للمعلمة (δ)

$-16212.736 \le \delta \le 46278.286$

أي ان التغير من موظف ليست له خبرة سابقة الى موظف له خبرة سابقة سيزيد راتبه السنوي ما بين (-16212.736) – (46278.286) ديناراً سنوياً.

ولبيانات المثال نفسها فان النتائج باستخدام OLS وبافتراض وجود تداخل بين المتغير النوعي والمتغير الكمي:

$$\hat{Y} = 5530.750 + 24.968X + 12976.137D + 330.378(XD)$$

اختبار خطي الانحدار متطابقان:

$$H_0 : \beta_2 = \beta_3 = 0$$

$$vs. \quad H_1 : \beta_2 \ne \beta_3 \ne 0$$

ان احدهما على الأقل لا يسأوي صفراً

$$F^* = \frac{ESS(D, XD/X)/2}{RSS(X, D, XD)/(n-k-1)} = \frac{ESS(X, D, XD) - ESS(X)/2}{RSS(X, D, XD)/16} = 0.912$$

$$F_{c(2,16,0.95)} = 3.63$$

بما ان $F^* > F_c$ فيكون خطا الانحدار متطابقين.

ولاختبار فيما اذا كان خطا الانحدار لهما معامل الانحدار نفسه.

أي لاختبار فيما اذا لم يوجد تداخل بينهما نتبع:

$$H_0 : \gamma = 0$$
$$vs. \quad H_1 : \gamma \neq 0$$

$$F^* = \frac{MES(XD/X, D)}{MRS(X, D, XD)} = 0.017 \ \langle \ F_{c(0.95)} = 3.63$$

$$ESS(XD/X, D) = ESS(X, D, XD) - ESS(X, D)$$

اذن لا يوجد تداخل بين خطي الانحدار.

أي ان معادلة التقدير: $\hat{Y} = 4423.144 + 206.213X + 15032.775D$

هي التي تمثل البيانات خير تمثيل.

مثال(10-14): دراسة دالة الاجر لعينة من 20 موظفاً بدلالة سنوات الخدمة X والحالة التعليمية للعينة وتتضمن ثلاث فئات: (ثانوية ، جامعة، دراسات عليا).

$$D_1 = \begin{cases} 1 & \text{جامعة} \\ 0 & \end{cases} \quad , \quad D_2 = \begin{cases} 1 & \text{عليا} \\ 0 & \end{cases}$$

Y	(الحلة التعليمية)			X_1 (متغير كمي)	رقم الموظف
	D_2		D_1		
3250	0	كلية	١	5	1
1600	1	عليا	٠	1	2
7500	٠	ثانوية	٠	16	3
13750	0	كلية	١	10	4
6000	0	ثانوية	0	5	5
8700	1	عليا	0	1	6
11350	0	كلية	1	3	7
3000	0	كلية	1	4	8
15700	1	عليا	0	15	9
113500	1	عليا	0	6	10
١٢٢٥٠	1	عليا	0	8	11
700	0	ثانوية	0	2	12
10250	1	عليا	0	4	13
3500	0	ثانوية	0	8	14
4500	0	ثانوية	0	10	15
16350	0	كلية	1	13	16
3800	0	كلية	1	6	17
9800	1	عليا	0	3	18
10800	0	كلية	1	2	19
23300	0	كلية	1	3	20

وباستخدام OLS اعطت النتائج التالية:

$$\hat{Y} = -982.238 + 661.249X + 7880.059D_1 + 21935.460D_2$$
$$t: \quad (0.514) \quad (0.551) \quad (1.486)$$
$$s.e: \quad 1285.579 \quad 14295.797 \quad 14758.984$$

ان عدم معنوية المعلمات δ_2 ، δ_1 تدل على ان خطوط الانحدار متطابقة.

$\hat{\beta}_1 = 661.249$: تعني ان زيادة سنة واحدة في الخدمة ستزيد الدخل السنوي مقدار 661.249 ديناراً (مع بقاء D_1 و D_2 ثابتة).

δ_1: تعني الفرق في الراتب السنوي لموظف حاصل على شهادة بكالوريوس عن موظف بشهادة ثانوية مقدار 7880.059 ديناراً.

أما الفرق بين الراتب السنوي لموظف حاصل على شهادة عليا عن موظف بشهادة ثانوية مقدار 21935.460 ديناراً، وعن موظف بشهادة بكالوريوس بمقدار 14055.401 ديناراً.

وبافتراض وجود تداخل بين المتغير الكمي والمتغيرات النوعية باستخدام طريقة المربعات الصغرى، اعطت النتائج التالية:

$$\hat{Y} = 1360.638 + 375.532X + 7425.473D_1 + 17125.6361D_2 - 42.682(XD_1) + 740.154(XD_2)$$

– اختبار فيما اذا كانت خطوط الانحدار متطابقة:

$$H_0: \delta_1 = \delta_2 = \gamma_1 = \gamma_2 = 0$$
$$vs. \quad H_1: \delta_1 \neq \delta_2 \neq \gamma_1 \neq \gamma_2 \neq 0$$

$$F^* = \frac{MES(D_1, D_2, XD_1, XD_2 / X)}{MRS(X, D_1, D_2, XD_1, XD_2)}$$

$$= \frac{ESS(X,, D_1, D_2, XD_1, XD_2) - ESS(X)/4}{RSS/14} = 0.55$$

$$F_{c(2,14,0.95)} = 3.11$$

$F^* < F_c$

أي ان خطوط الانحدار متطابقة.

ولاختبار معنوية التداخل بين المتغير النوعي والكمي

$$H_0: \gamma_1 = \gamma_2 = 0$$
$$vs. \quad H_1: \gamma_1 \neq \gamma_2 \neq 0$$

$$F^* = \frac{MES(XD_1, XD_2 / D_1, D_2, X)}{MRS(X, D_1, D_2, XD_1, XD_2)} = \frac{ESS(X, D_1, D_2, XD_1, XD_2) - ESS(X, D_1, D_2)/2}{RSS(X, D_1, D_2, XD_1, XD_2)/14}$$
$$= 0.037$$

$$F^* \langle \ F_{c(2,14,0.95)}$$

اذن لا يوجد تداخل بين المتغير الكمي والمتغير النوعي.

أي ان معادلة التقدير السابقة هي أفضل تمثيل للحالة

$$\hat{Y} = -982.238 + 661.249X + 7880.059D_1 + 21935.460D_2$$

مثال(13-10): دراسة العلاقة بين الراتب السنوي (Y) وعدد سنوات الخدمة (X) والخبرة الادارية السابقة (D_1) والحالة التعليمية.

علماً ان D_1 هو متغير نوعي بفئتين

وان الحالة التعليمية وهي متغير نوعي له ثلاث فئات. يمثل بواسطة متغيرين D_2 ، D_3.

$$D_1 = \begin{cases} 1 & \text{له خبرة} \\ 0 & \text{ليست له خبرة} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{شهادة كلية} \\ 0 & \text{غير ذلك} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{شهادة عليا} \\ 0 & \text{غير ذلك} \end{cases}$$

اعطيت معلومات حول:

دالة الاستجابة بفرض عدم وجود تداخل: $\mathrm{E}(Y) = \hat{\beta}_0 + \hat{\beta}_1 X + \delta_1 D_1 + [\delta_2 D_2 + \delta_3 D_3]$

وعليه فان دالة الاستجابة للحالات المختلفة كالاتي:

$$E(Y)=\hat{\beta}_0+\hat{\beta}_1X$$

موظف ليست له خبرة وحاصل على شهادة ثانوية.

$$E(Y)=(\hat{\beta}_0+\hat{\delta}_1)+\hat{\beta}_1X$$

موظف له خبرة وحاصل على شهادة ثانوية.

$$E(Y)=(\hat{\beta}_0+\hat{\delta}_2)+\hat{\beta}_1X$$

$$E(Y)=(\hat{\beta}_0+\hat{\delta}_1+\hat{\delta}_2)+\hat{\beta}_1X$$

موظف ليست له خبرة وخريج جامعة.

$$E(Y)=(\hat{\beta}_0+\hat{\delta}_3)+\hat{\beta}_1X$$

$$E(Y)=(\hat{\beta}_0+\hat{\delta}_1+\hat{\delta}_3)+\hat{\beta}_1X$$

موظف له خبرة وخريج جامعة

δ_1: تقيس مقدار الفرق بالراتب السنوي بسبب الخبرة التي يحصل عليها خريج الثانوية.

δ_2: تقيس مقدار الفرق بالراتب السنوي لخريج الكلية مقارنة لشهادة الثانوية.

δ_3: تقيس مقدار الفرق بالراتب السنوي بسبب الشهادة العليا نسبة الى الشهادة الثانوية.

$(\delta_3 - \delta_2)$: تقيس مقدار الفرق بالراتب السنوي للشهادة العليا نسبة الى الشهادة الكلية.

كما ان معادلة التقدير:

$$\hat{Y}=1395.550+534.722X+2768.826D_1+2224.934D_2+7254.145D_3$$

$$t: \quad (10.53)^{**} \quad (4.75)^{**} \quad (3.29)^{**} \quad (14.36)^{**}$$

وتشير قيم t الى معنوية المعلمات المقدرة.

$\hat{\beta}_1$: تشير الى ان الراتب السنوي سيزداد بمقدار (534.722) ديناراً لكل سنة زيادة في الخدمة مع بقاء بقية المتغيرات ثابتة.

δ_1: تشير الى ان مقدار الفرق بالراتب السنوي بسبب الخبرة الادارية هو 2768.826 ديناراً.

δ_2: تشير الى ان مقدار الفرق بالراتب السنوي لخريج الكلية عن خريج الثانوية: 2224.934 ديناراً.

δ_3: تشير بان مقدار الفرق بالراتب السنوي العليا عن خريج الثانوية: 7254.145 ديناراً.

أما الفرق بالراتب السنوي لموظف حاصل على شهادة عليا وآخر خريج كلية فهو: 5029.211 ديناراً.

وفي حالة وجود تداخل بين المتغيرات المستقلة فان دالة الاستجابة تتمثل كالآتي:

$$E(Y)=\hat{\beta}_0+\hat{\beta}_1X+\hat{\delta}_1D_1+\hat{\delta}_2D_2+\hat{\delta}_3D_3+\hat{\delta}_4(XD_1)+\hat{\delta}_5(XD_2)+\hat{\delta}_6(XD_3)+\hat{\delta}_7(D_1D_2)+\hat{\delta}_8(D_1D_3)$$

اسئلة الفصل العاشر

س1: بيانات عن أسعار السكن (Y) ومساحة السكن (X_1)، عمر بناء السكن (X_2) لـ (4682) سكناً تم بيعها في محافظة معينة للسنوات 1991 كانون الثاني إلى 1996 كانون الأول.
لدراسة علاقة الانحدار لتحديد أسعار السكن في تلك المحافظة تتم إضافة متغيرات وهمية لسنوات الدراسة كآلاتي:

$$D_1 = \begin{cases} 1 & \text{السكن تم بيعه في 1992} \\ 0 & \text{غير ذلك} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{السكن تم بيعه في 1993} \\ 0 & \text{غير ذلك} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{السكن تم بيعه في 1994} \\ 0 & \text{غير ذلك} \end{cases}$$

$$D_4 = \begin{cases} 1 & \text{السكن تم بيعه في 1995} \\ 0 & \text{غير ذلك} \end{cases}$$

$$D_5 = \begin{cases} 1 & \text{السكن تم بيعه في 1996} \\ 0 & \text{غير ذلك} \end{cases}$$

وبذلك تكون معادلة الانحدار:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \delta_1 D_1 + \delta_2 D_2 + \delta_3 D_3 + \delta_4 D_4 + \delta_5 D_5$$

(1): فسر معلمات المتغيرات الوهمية المقدرة. اذا تم استخدام الصيغة الخطية.

(2): فسر معلمات المتغيرات الوهمية المقدرة. اذا تم استخدام الصيغة اللوغاريتمية المزدوجة.

(٣) :وضح كيف تختبر أنه لاتأثير لسنة البيع على السعر .

س2: نموذج انحدار لدراسة الدخل الشخصي السنوي (Y) كدالة بدلالة سنوات الدراسة (E) والتحصيل العلمي لعينة تحصيلهم الدراسي (بكالوريوس ، اعدادية ، ابتدائية).

م // (1): وصف كيف يتم تضمين متغير التحصيل العلمي بالانحدار.

(2): اذا علمت ان معادلة التقدير : $\hat{Y} = 4.5 + 0.5E - 1.0D_1 + 1.5D_2$

ارسم الدخل المقدر كدالة بدلالة سنوات الدراسة X وحسب التحصيل العلمي للمستويات الثلاث علماً ان الدراسة للعينة المختارة هي ما بين (8-16).

(3) اذا علمت ان معادلة التقدير بوجود تداخل بين E والمتغيرات الوهمية هي

$$\hat{Y} = 4.5 + 0.5E + 1.4D_1 + 0.3D_2 - 0.2D_1 + 0.1D_2$$

ارسم خط انحدار الدخل المقدر ولمستويات التحصيل العلمي الثلاث.

س3: البيانات التالية عن المشتريات من السندات الحكومية B والدخل القومي Y للسنوات(1974-1990)

B	2.6	3.0	3.6	3.7	3.8	4.1	4.4	7.1	8.0	8.9	9.7	10.2	10.1	7.9	8.9	9.1	10.1
Y	2.4	2.8	3.1	3.4	3.9	4.0	4.2	5.1	6.3	8.1	8.8	9.6	9.7	9.6	10.4	12.0	12.9

م/(1): رسم الانتشار.

(2): هل ان رسم الانتشار يوضح ان هناك نمطين من مشتريات السندات، اذا كان الجواب بالإيجاب، اقترح حلاً ملائماً لتقدير العلاقة بين المشتريات من السندات الحكومية والدخل القومي.

المصـــادر :

١– الراوي، خاشع محمد،١٩٨٧. " المدخل الى تحليل الانحدار "، دار الكتب للطباعة والنشر، الموصل.

٢– شربجي،عبد الرزاق محمد صلاح، ١٩٨١. " الانحدار الخطي المتعدد "، دار الكتب للطباعة والنشر، الموصل.

٣– عبد الرحمن، عبد المحمود محمد " مقدمة في الاقتصاد القياسي "، جامعة الملك سعود للطباعة والنشر.

٤– عطية، عبد القادر محمد عبد القادر، 2004. " الحديث في الاقتصاد القياسي: بين النظرية والتطبيق"،مكة المكرمة.

٥– كلجيان، هاري و اوئتس، والس، 1995. " مقدمة في الاقتصاد القياسي المبادئ والتطبيقات "، ترجمة المرسي السيد حجازي وعبد القادر محمد عطية، جامعة الملك سعود.

٦– نتر، جون ، وازمان ويليام و كتنر، ميخائيل " نماذج احصائية تطبيقية" ، ترجمة انيس اسماعيل كنجد، عبد الحميد بن عبد الله الزيد وابراهيم بن عبد العزيز الواصل والحسيني عبد البر راضي، الجزء الاول (الانحدار) جامعة الملك سعود للطباعة والنشر.

٧– هادي، اموري و مسلم، باسم شلبية، 2002. " القياس الاقتصادي المتقدم ، النظرية والتطبيق"، مطبعة الطيف.

1- Draper, N.R. & Smith, H., 1966. " Applied Regression Analysis ", New York, john Wiley & Sons , Inc.

2- Franses, Philip Hans , " A concise introduction to econometrics ", An intuitive guide, Cambridge . org .

3- Gujarati, Damodar N. 1995." Basic Econometrics " , 3rd ed., ., Me Grow – Hill international, Edition, Singapore

4- Johnston, J., 1984. " Econometric methods ", 3rd ed., Me Grow – Hill international, New York.

5- Maddala, G.S., 2002." Introduction to Econometrics ", Third Edition, , New York, John willey & sons, 1 td .

جملة من منشورات عبر الويب.

View publication stats